RECHERCHE EN COMPORTEMENT

SUR TELIDON I

Groupe de Recherche en Comportement

Ministère des Communications

1980

RECHERCHE EN COMPORTEMENT AXEE SUR LE TELIDON I
MINISTERE DES COMMUNICATIONS
OTTAWA
SEPTEMBRE 1979

Préface

Le Télidon, le système vidéotex mis au point au Canada, sera mis à l'essai dans les foyers canadiens, dans le cadre d'expériences-pilotes dont les préparatifs sont en cours. De nombreuses décisions d'ordre fonctionnel qui seront prises pendant cette étape préparatoire auront des répercussions sur les utilisateurs. Le Groupe de recherche en comportement du ministère des Communications a pesé plusieurs de ces décisions et a mené des études tant pour dégager celles qui sont susceptibles d'influer sur les utilisateurs que pour aider à formuler des recommandations à l'intention des concepteurs. La série de rapports présentée ici résume les résultats des travaux réalisés jusqu'à maintenant. C'est le premier document ainsi présenté. Il se veut un instrument de communication avec ceux qui préparent comme nous les essais du Télidon.

Les rapports présentés dans le document se répartissent en trois catégories: exposés de facteurs humains mis en cause par la technologie du Télidon, études de caractéristiques d'affichage visuel convenant le mieux aux utilisateurs, études de caractéristiques des bases de données convenant le mieux aux utilisateurs.

1. Exposés de facteurs humains mis en cause.

Le premier document qui s'intitule le Télidon, le système vidéotex canadien et ses répercussions sur les sciences de l'information présente un premier ensemble de questions portant sur la création de l'information dans le système Télidon et examine les possibilités et les limites du Télidon sur le plan technique.

Le système vidéotex Télidon et les questions se rapportant aux utilisateurs aborde, en rapport avec les utilisateurs, les questions concernant l'affichage visuel et le protocole de dialogue pour fins de recherche documentaire.

Le document consacré aux répercussions du Télidon sur les sciences de l'information présente un modèle général permettant de comprendre la recherche dont ce système fait l'objet et fait le point sur les problèmes que soulèvent la planification et l'organisation d'une banque de données.

2. Facteurs humains rattachés à l'affichage visuel.

Quatre rapports portent sur des travaux de recherche menés au Centre de recherches sur les communications. Le document intitulé Jeux de caractères des terminaux vidéotex servant aux expériences-pilotes résume des études effectuées à titre d'expérimentation et de démonstration et il recommande l'espacement optimal des caractères suivant le double critère de l'économie et de la facilité de lecture.

Les deux exposés qui suivent soit Espacement des caractères sur un écran de télévision: (jeu de caractères 5 sur 7 et Espacement de caractères sur un écran de télévision: jeu de caractères 7 sur 9 présentent les résultats d'études empiriques portant sur la lisibilité des caractères selon l'espacement employé. Les auteurs concluent en recommandant les espacements qui conviennent le mieux à chaque jeu de caractères sur l'écran de visualisation du Télidon.

Accentuation des caractères est une étude des différentes techniques d'accentuation des caractères employés en français. La recommandation sur la méthode jugée la plus acceptable par les utilisateurs découle de données de préférence recueillies auprès de francophones.

3. Facteurs humains rattachés aux bases de données.

Deux exposés intitulés Nombre optimal de rubriques que devrait proposer une page d'index dans la base de données du système Télidon interactif et Le Télidon en mode de radiodiffusion: nombre optimal de rubriques par page d'index présentent un modèle dont devrait s'inspirer la conception de la base de données du système Télidon afin de permettre aux utilisateurs de trouver les renseignements voulus le plus rapidement possible. En tenant compte de plusieurs variables, le modèle recommande d'inclure entre 4 et 8 rubriques par page d'index dans le cas du Télidon interactif (voir les détails dans le résumé de chacun des articles). Dans le cas du Télidon en mode de radiodiffusion, tant pour la transmission sur la totalité d'un canal que pour la transmission pendant les intervalles de suppression verticale, une dizaine de rubriques par page serait près du nombre optimal. Ces deux articles étudient d'autres recommandations (voir les détails dans le résumé de chacun des articles).

L'exposé intitulé Évaluation de la structure arborescente de l'information dans la base de données du Télidon vise à évaluer la difficulté qu'auront les utilisateurs à consulter la base de données. Dans l'ensemble, il s'agit d'une première tentative d'élaboration d'une méthode d'évaluation et d'amélioration des bases de données à structure arborescente.

Quelques-uns des rapports inclus dans le présent document ont déjà été publiés tel qu'il est indiqué dans les exposés. Les commentaires et questions portant sur ce programme de recherches doivent être transmis aux auteurs du document à l'adresse suivante:
300 rue Slater, Ottawa (Ontario), Canada, K1A 0C8.

TABLE DES MATIERES

Préface

D. A. Phillips ii

I Exposés de facteurs humains mis en cause. 1

1. Télidon, le système vidéotex canadien et ses répercussions sur les sciences de l'information. 2

D.A. Phillips

2. Le système vidéotex Télidon et les questions se rapportant aux utilisateurs. 10

H. Bown, C.D. O'Brien W. Sawchuk, J.R. Storey et W. Treurniet

3. Répercussions de Télidon dans le monde de l'information 28

D. A. Phillips

II Facteurs humains rattachés à l'affichage visuel. 44

4. Espacement des caractères sur un écran de télévision (Jeu de caractères 5 sur 7). 45

W. C. Treurniet

5. Jeux de caractères des terminaux vidéotex servant aux expériences - pilotes. 71

W. C. Treurniet et C. D. O'Brien

6. Espacement des caractères sur un écran de télévision (Jeu de caractères 7 sur 9). 106

W. C. Treurniet

7. Accentuation des caractères. 117

W. C. Treurniet et S. Latrémouille

III Facteurs humains rattachés aux bases de données. 125

8. Nombre optimal de rubriques que devrait proposer une page d'index dans la base de données du système Télidon interactif. 126

E. Lee

9. Le Télidon en mode de radiodiffusion: nombre optimal de rubriques par page d'index. 150

E. Lee

10. Evaluation de la structure arborescente de l'information dans la base de données du Télidon. 228

E. Lee et S. Latrémouille

CHAPITRE I

EXPOSES DE FACTEURS HUMAINS MIS EN CAUSE

TELIDON, LE SYSTEME VIDEOTEX CANADIEN ET SES REPERCUSSIONS SUR LES SCIENCES DE L'INFORMATION

Dorothy A. Phillips
Groupe de recherche en comportement
Recherche et développement, techniques et systèmes
Ottawa (Ontario)
K1A 0C8

RESUME

Télidon est le nom donné au système vidéotex canadien, un service interactif de recherche documentaire à accès public qui est en voie d'élaboration à Ottawa. A l'échelle internationale, les systèmes vidéotex et télétext (radiodiffusion) font l'objet de plusieurs autres réalisations grâce auxquelles l'information graphique et alphanumérique est accessible aux entreprises et aux usagers privés disposant d'écrans de visualisation (téléviseurs). Les caractéristiques techniques de l'affichage et de la structure de recherche du Télidon ont des répercussions sur la façon d'y présenter l'information. D'autre part, des questions demeurent en suspens quant à la nature de l'information se prêtant le mieux au Télidon, aux schémas d'indexation ou de classification et aux droits à exiger en retour de l'information. Le monde des sciences de l'information est invité à faire valoir ses idées sur le sujet.

Extrait des Actes de la réunion annuelle de l'Association canadienne des sciences de l'information, 1979.

INTRODUCTION

Dans le monde des sciences de l'information, on n'est pas sans savoir que les progrès technologiques sont en train d'y créer une véritable révolution. Le fusionnement actuel des nouvelles techniques exploitées en informatique et en télécommunications modifiera profondément notre façon d'obtenir et de stocker l'information. Il est permis de croire que, d'ici quelques années, quiconque possèdera un appareil téléphonique ou un téléviseur, aura accès à des bases de données mécanographiques. Nous pourrons dialoguer avec la base de données pour connaître les résultats des dernières joutes de hockey ou le cours du marché, choisir une nouvelle maison en examinant des plans et des illustrations, obtenir des renseignements sur les conditions d'adoption d'un enfant ou encore l'adresse d'un conseiller en orientation. Ces services de recherche documentaire sont présentement en voie d'élaboration. Plus tard, le même matériel jouera le rôle de messager. Il nous permettra d'écrire à nos amis, de réserver des places pour le théâtre ou un voyage en avion, de faire le marché et d'en acquitter la facture et peut-être même d'être interrogés par l'ordinateur sur nos ennuis de santé avant d'aller chez le médecin.

Le Télidon est une des réalisations technologiques qui sont appelées à avoir des répercussions importantes sur nos vies dans un avenir rapproché. Télidon est le nom qui a été donné au système vidéotex canadien, vidéotex étant le terme générique adopté temporairement dans le monde entier pour désigner les services interactifs de recherche documentaire à accès public. Le Télidon a été mis au point au Centre de recherche sur les communications à Ottawa.

Le présent exposé décrit d'abord quelques-unes des caractéristiques du Télidon et situe ce dernier dans le contexte des autres réalisations internationales dans ce domaine. Il réclame ensuite la participation des praticiens de l'information à l'élaboration de l'information qui sera diffusée au moyen du Télidon.

VIDÉOTEX ET TÉLÉTEXT

Avant de décrire le Télidon en détail, il est utile de faire la distinction entre les services vidéotex bilatéraux et les services télétext unilatéraux. Les systèmes vidéotex permettent à l'usager d'accéder à l'information stockée en mémoire à l'aide d'un clavier numérique ou d'un clavier plus complexe, d'un téléviseur modifié ou d'un autre écran de visualisation et d'une ligne de transmission téléphonique ou autre. Quant aux services télétext unilatéraux, ils sont en mode de radiodiffusion et l'usager peut au moyen d'un décodeur et d'un téléviseur "attraper" une image au moment où elle passe dans le signal radiodiffusé et l'afficher sur l'écran de son téléviseur. Dans le cas du télétext, le temps d'accès dépend du temps nécessaire au déroulement de la page et peut atteindre une minute selon l'importance de la base de données.

Les systèmes télétext comme les systèmes vidéotex affichent des caractères alphanumériques et des graphiques en couleur. Du point de vue de l'usager, les systèmes vidéotex peuvent offrir un accès à des bases de données beaucoup plus importantes et le temps de réponse est plus court en moyenne.

Les services télétext sont présentement offerts en Grande-Bretagne par la BBC (Ceefax) et par ITV (Oracle).

Dans le cas des systèmes Ceefax et Oracle, les organismes de radiodiffusion fournissent et les services de mise en forme ou de cueillette des données et les voies de diffusion, ce qui diffère sensiblement des systèmes vidéotex, pour lesquels les services d'information et de télécommunication ont été fournis jusqu'à aujourd'hui par des groupes différents. (Pye, 1978).

Réalisations internationales concernant le vidéotex

Le Télidon, réalisé au Canada, s'ajoute à d'autres systèmes bilatéraux de recherche documentaire du même genre mis au point ailleurs dans le monde. Les Postes britanniques ont élaboré le système Prestel (nommé à l'origine Viewdata) qui, après un essai de commercialisation auprès d'une clientèle expérimentale de 1600 usagers, devrait être offert au public à compter du printemps de 1979. On espère avoir à ce moment-là 250 000 pages d'information mises à la disposition du public par tout près de 200 fournisseurs d'information indépendants.

Le système français Titan, un service vidéotex interactif, utilise un système de visualisation nommé Antiope mis au point dans des laboratoires français. Encore au stade expérimental en laboratoire, le système Titan doit être mis à l'essai sur le terrain en 1980.

Ailleurs en Europe, les organismes téléphoniques de l'Allemagne de l'Est et des Pays-Bas ont acheté du matériel Prestel et d'autres pays paraissent s'y intéresser. Plusieurs sociétés américaines s'intéressent également à la technologie du Prestel. Le Japon, pour sa part, a réalisé plusieurs expériences dans le domaine du vidéotex et du télétext en utilisant sa propre technologie.

A mesure que s'élaborent les plans des essais sur le terrain auxquels doit être soumis le système vidéotex canadien Télidon, l'intérêt grandit sur la scène internationale. En 1979, à la mi-février, Bell Canada lançait en collaboration avec les sociétés Torstar et Southam Press une expérience-pilote de son propre système vidéotex connu sous le nom de Vista, dans le cadre de laquelle le Télidon sera aussi mis à l'essai. D'autres expériences-pilotes doivent être entreprises de concert avec le ministère des Communications par le Manitoba Telephone System, l'Alberta Government Telephones, l'Office de la télécommunication éducative de l'Ontario, Télécâble Vidéotron à Montréal et Grand River Cable pour mettre à l'essai le système télétext.

ASPECTS TECHNIQUES DE LA FOURNITURE D'INFORMATIONS AVEC LE TELIDON

Caractéristiques de l'affichage

Dans le cas du Télidon, le processus de création et d'affichage des pages d'information est tributaire du mode de stockage et de transmission des caractères et des graphiques qui diffère des systèmes britanniques ou français. Ces systèmes européens transmettent des images sous forme d'éléments visuels mosaïques. L'image visuelle du téléviseur britannique à 625 lignes se compose de 24 lignes de 40 caractères et celle du système français de 25 lignes de 40 caractères. Les images graphiques sont constituées de caractères graphiques en forme de blocs qui forment les éléments d'une mosaïque. Le système Télidon, par contre, a été mis au point suivant un jeu d'instructions de description de de l'image (IDI) qui décrivent l'image en termes d'éléments goéométriques

primaires (points, lignes, cercles, rectangles, polygones, arcs), d'images semblables aux fac-similés exécutées bit par bit et de textes. Ce concept permet la création graduelle de vastes bases de données qu'il ne sera pas nécessaire de modifier en raison du perfectionnement des terminaux ou de l'adoption de nouveaux moyens de transmission. (Bown et al., 1978)

Pour le moment, le format du téléviseur nord-américain à 525 lignes restreint la présentation des textes par Télidon à 20 lignes de 40 caractères, mais l'avènement de visuels à résolution plus fine fera disparaître cette restriction. Présentement, la gamme de couleurs qu'il est possible de transmettre sur le Télidon se limite à 8, y compris le noir et le blanc, à laquelle s'ajoute une échelle de 6 degrés de gris.

Base de données à structure arborescente

Présentement, seules les techniques de recherche par arborescences permettent l'accès aux bases de données. Au départ, c'est l'économie de temps de traitement qui a inspiré la mise au point des structures arborescentes du Télidon à l'instar de celles du Prestel. On prévoit que beaucoup plus de gens utiliseront simultanément le service Télidon que même la plus populaire des bases de données actuelles. L'économie des ressources informatiques est donc importante afin de conserver le temps de réponse à un niveau raisonnable.

La structure arborescente du Télidon comporte une page d'index général, la page 0, suivie de 9 autres niveaux de pages. Chaque niveau offre 9 possibilités. On peut aussi avoir accès à une liste séquentielle de 1000 pages sur n'importe quelle ramification de la structure. Il y a au total 10^{11} adresses possibles, ce qui, de toute évidence, est énorme. Comme le système britannique, après plus d'une année, atteint à peine 2×10^5 pages, nous n'avons pas à nous sentir limités par l'espace. Rien n'empêche, non plus, l'existence de plus d'une base de données.

Chaque page de la structure porte une adresse unique. L'utilisateur peut y avoir accès en passant par les ramifications de la structure ou en la choisissant directement. Les adresses peuvent être répertoriées alphabétiquement sur une des pages de la structure arborescente. D'autre part, il est possible de revenir à la page 0 à partir de n'importe quelle page et de recommencer sa recherche dans la structure, ainsi que de reculer et d'avancer d'une page ou plus à la fois dans la base de données.

La technologie du Télidon assure un potentiel d'affichage de graphiques et de textes auquel seule l'imagination du concepteur de pages pourrait mettre un frein. La structure arborescente pourvoit à une très grande base de données, mais, pour le moment, elle limite la méthode de recherche au choix d'un article parmi neuf possibilités à chaque niveau.

QUESTIONS QUE SOULEVE LA FOURNITURE DE L'INFORMATION

Les facteurs technologiques qui déterminent le champ des possibilités de visualisation et d'extraction de l'information qu'offre le Télidon soulèvent un certain nombre de questions auxquelles les praticiens de l'information pourraient apporter des éléments de réponse. Ces questions sont en fait un défi que je lance aux scientifiques canadiens du monde de l'information. C'est d'eux que

devront venir les réalisations techniques nécessaires et la base rationnelle des décisions qui détermineront comment le Télidon nous sera présenté dans nos maisons.

Ces questions peuvent être abordées sur deux plans: "Que pouvons-nous faire avec la technologie dans son état actuel?" et, du point de vue de l'utilisateur, "Quel ordre de préséance faut-il accorder aux perfectionnements techniques qui restent à réaliser? Lesquels auront le plus d'importance pour l'utilisateur?"

Les questions auxquelles vous êtes appelés à répondre sont donc les suivantes:

1. A la présentation de quelles informations la technologie du Télidon se prête-elle le mieux?
2. Quelle(s) méthode(s) d'indexation doit-on adopter pour la composition des pages-parapluie ou pages d'acheminement de la structure arborescente? Une normalisation est-elle possible?
3. De quelle façon les utilisateurs devraient-ils payer l'information qu'ils reçoivent au moyen du Télidon?

Contenu de l'information

L'information nécessaire à la prestation des services Télidon aux entreprises et aux usagers privés sera fournie par des organismes ou des personnes désignés sous le nom de fournisseurs d'information. Dans le cas de Prestel, les Postes britanniques se sont chargées de fournir la voie de transmission, mais non le contenu qui vient de fournisseurs d'information. Ces derniers, dont le volume de production va de modeste à considérable, peuvent stocker jusqu'à 10 000 pages dans la base de données.

Au Canada, la transmission des communications ne sera pas assumée par l'État, mais on n'a pas encore établi clairement si cette responsabilité sera confiée aux compagnies de téléphone ou de câbles ou aux deux. On n'a pas non plus déterminé à qui seront dévolues les fonctions de fournisseurs d'information. Une des principales questions des prochaines années portera sur la nature de l'information à présenter sur Télidon. A quels renseignements l'utilisateur voudra-t-il avoir accès sur son écran Télidon? Quelle sorte de renseignements utilisera au mieux les possibilités graphiques du Télidon et la facilité d'accéder instantanément à des informations en évolution rapide. Quels renseignements l'utilisateur acceptera-t-il de payer? Certains renseignements devraient-ils être fournis gratuitement? Le Télidon servira-t-il de substitut ou de complément aux journaux et à d'autres publications? Les seules données empiriques dont nous disposons nous viennent de la Grande-Bretagne où un périodique intitulé Prestel User's Guide a été mis sur le marché. Loin de remplacer les publications existantes, Prestel en a engendré une. Prestel fournit une gamme variée de renseignements: nouvelles, des horaires de trains, renseignements sur les restaurants, les voyages, le cinéma, les assurances et nombre d'autres sujets.

Les essais-études de marché effectués en Grande-Bretagne ne nous ont pas encore livré de données sur le genre de renseignements que préfèrent les utilisateurs britanniques. Dès que nous en aurons, nous les étudierons avec grand soin même si les désirs du public canadien ne correspondent pas nécessairement à ceux du public britannique et si le système Télidon, avec ses différences techniques, permet de présenter des informations différentes.

Je crois que les praticiens de l'information, qui connaissent, pour y avoir pourvu notamment dans les bibliothèques, les besoins et les désirs du public en matière de documentation écrite, sont à même de contribuer au contenu informationnel du Télidon. L'industrie de l'édition a déjà manifesté de l'intérêt à l'égard du système. La presse (le Globe and Mail, Southam, Torstar et la Presse) s'intéresse activement aux essais qui s'y rapportent. Beaucoup d'autres créateurs d'information, tant privés que publics, pourraient contribuer au contenu du Télidon.

Indexation de l'information

La recherche par arborescences qui permet présentement d'accéder au Télidon suscite une deuxième série de questions. Etant donné les limites de cette méthode de recherche, de quelle façon les informations devraient-elles être répertoriées afin que l'utilisateur puisse y accéder facilement? Si diverses banques de données destinées au Télidon sont mises au point dans tout le pays, comme tout semble l'indiquer, il paraîtrait raisonnable d'établir des index normalisés de sorte qu'un utilisateur itinérant puisse trouver ce qu'il cherche où qu'il soit dans le pays. Est-il possible de mettre au point un index ou un jeu d'index qui servirait de norme? Y a-t-il lieu d'entreprendre une activité de normalisation de ce genre dans l'immédiat?

L'indexation de l'information peut prendre une importance toute particulière si elle représente un instrument de filtrage pour l'utilisateur. Quand ce dernier choisit une ramification donnée de la structure, il ne peut être exposé qu'à un renseignement précis, processus bien différent de celui qui consiste à feuilleter un périodique. Les gens seront-ils enclins à ce genre de filtrage alors qu'ils se dirigent vers une saturation de l'information. (Conseil des sciences du Canada, octobre 1978). Est-il possible de concevoir des systèmes de renvoi ou d'indexation qui éviteront les effets négatifs possibles du choix que devra faire l'utilisateur? A ma connaissance, personne n'a encore abordé ces questions d'indexation.

Si nous projetions nos pensées encore plus loin dans l'avenir, quelles structures de recherche seraient le mieux adaptées aux utilisateurs si la technologie atteignait un point où les restrictions qu'impose la structure arborescente étaient éliminées? A quelles informations les utilisateurs pourraient-ils avoir accès au moyen de structures différentes? Quelle importance celà représenterait-il pour eux?

Une autre série de questions qui se posera dans l'avenir concerne les langages d'interrogation et leur acceptabilité par le public. Dialoguer avec une base de données à structure arborescente est très simple; il suffit de faire une décision parmi des choix multiples et de pousser quelques touches d'un clavier numérique. Si un système conversationnel plus complexe était mis au point et des claviers complets installés dans toutes les maisons, quel langage d'interrogation serait le mieux adapté au grand public?

Paiement de l'information par l'utilisateur du Télidon

L'information n'est pas gratuite. Même l'information offerte par les bibliothèques publiques est payée par nos impôts. Néanmoins, nous vivons dans une société dont une des valeurs consiste à donner à tout individu, quels que soient ses moyens financiers, libre accès à l'information. Il est évident que tous les renseignements ne sont pas divulgués, mais au moins la plus grande partie des archives importantes de notre société est accessible au public. Quel mode de paiement doit-on, dans ce cas, adopter en retour de l'information offerte par le Télidon?

En Grande-Bretagne, on avait d'abord opté, dans le cas du Prestel, pour un prix par page d'information. Ce système a cependant été modifié récemment de sorte que les utilisateurs ne paient maintenant que les pages finales qui contiennent l'information recherchée, et non les pages d'index. Par ailleurs, des frais supplémentaires s'appliquent à la durée de la communication, ces frais diminuant après 18h.

Si les pages d'index du Télidon sont gratuites, doit-on fixer un prix par page ou par minute ou doit-on opter pour un autre mode de paiement?

Le Conseil des sciences du Canada a soulevé une question connexe (décembre 1978) "L'information est-elle un droit ou un produit de consommation?" En choisissant le mode de paiement par page ou par minute, nous traitons l'information comme un produit de consommation. Ce choix aura peut-être aussi des répercussions sur le genre d'information que nous jugerons opportun de diffuser par Télidon. Si, par contre, l'information est payée par nos impôts et si le public y a accès gratuitement, l'information est traitée comme un droit et la valeur que nous attachons au libre accès de tous les citoyens à l'information se trouve confirmée. Cependant, il est possible qu'en procédant ainsi nous affaiblissions un secteur important de notre économie, les fournisseurs d'information, qui tirent un profit de la vente d'information.

Aucune de ces questions n'a été réglée. Dans certains cas, des progrès techniques sont nécessaires; des méthodes d'indexation et peut-être des modes d'interrogation en langage naturel peuvent être mis au point. On doit, dans certains cas, faire une analyse raisonnée fondée sur l'expérience que nous ont permis d'acquérir nos recherches antérieures sur les médias et le marché. Quant au genre d'information à présenter au moyen du Télidon, on peut y aller par tâtonnement et raffiner nos concepts à la suite d'études de marché et d'essais. Certaines questions doivent non seulement être soumises à une analyse raisonnée des possibilités, mais également être débattues. On peut délibérer sur la question du paiement qu'on exigerait de l'usager en retour de l'information ou on peut laisser le marché trancher la question. Toutes ces questions influeront sur la qualité de l'information à laquelle nous aurons accès dans un avenir rapproché grâce au Télidon.

DOCUMENTATION

Bown, H., O'Brien, C.D., Sawchuk, W. et Storey, J.R. Description générale du Télidon: Proposition canadienne concernant les systèmes vidéotex, Note technique n° 697F-CRC, ministère des Communications, Ottawa (Ontario), décembre 1978.

La télématique: information de la société canadienne. Un exposé de la situation, Conseil des sciences du Canada, octobre 1978.

La télématique: information de la société canadienne. Projet de programme d'étude. Conseil des sciences du Canada, décembre 1978.

Pye, R., Focus on Europe: Information Retrieval services Telecommunications Policy, March 1978.

Le système vidéotex Télidon et les questions se rapportant aux utilisateurs

par

H. Bown, C.D. O'Brien, W. Sawchuk, J.R. Storey et W.C. Treurniet

Centre de recherches sur les communications

C.P. 11490, succursale "H"

Ottawa (Ontario), Canada

K2H 8S2

Le présent exposé a été présenté à la
Conference on Processing of Visible Language
qui s'est tenue à Niagara-on-the-Lake le
4 septembre 1979 et sera versé aux actes
de la conférence.

Résumé

Les systèmes vidéotex sont des systèmes de communication visuelle interactifs destinés, à l'origine, à permettre au public d'accéder à des sources d'information automatisées. L'utilisateur peut choisir des pages d'information dans une base de données en transmettant des ordres au moyen d'un clavier numérique. La page commandée est affichée à l'écran d'un téléviseur (récepteur) ordinaire.

L'auteur expose ici le problème de l'adaptabilité des systèmes au regard de la description de l'image et il décrit une solution, sur laquelle repose le système canadien Télidon. Il aborde aussi, en rapport avec les utilisateurs, diverses questions concernant les modalités d'affichage de l'information et d'interface des programmes.

Vidéotex est le nom générique donné aux services de recherche documentaire qui font usage de récepteurs de télévision modifiés en guise de terminaux de visualisation. A l'aide d'un clavier, numérique ou autre, les utilisateurs du vidéotex pourront commander à l'ordinateur d'une base de données éloignée l'affichage d'une image particulière sur leur écran. L'image que permet d'afficher la technologie actuelle doit être essentiellement fixe; elle peut présenter l'information sous forme de texte, de dessins géométriques ou d'objets plus complexes ressemblant à des photographies. On entre en communication avec l'ordinateur au moyen de lignes téléphoniques ou d'autres réseaux interactifs. Le concept du vidéotex est révolutionnaire en ce sens qu'il donne désormais à l'utilisateur la possibilité d'exercer sur les images de son téléviseur un contrôle qui dépasse de beaucoup ce qui existait jusqu'ici.

Les pages d'information destinées à la base de données d'un système vidéotex sont créées au moyen de terminaux plus perfectionnés. L'information générée par un terminal de ce genre peut être destinée à des fins lucratives ou publicitaires. Par exemple, dans le premier cas, un écrivain peut vouloir employer le service pour publier un ouvrage de façon économique. Dans le second cas, un magasin peut y avoir recours pour annoncer sa marchandise. On pourra même un jour conclure des ventes à distance au moyen de ce système si on parvient à surmonter les obstacles qui empêchent encore les transferts de caisse électroniques.

Les progrès technologiques permettront sans doute aux terminaux de converser directement l'un avec l'autre sans l'intermédiaire d'un ordinateur

principal. Ainsi, lorsque les deux écrans répondront de façon identique aux directives originant de l'un ou l'autre des terminaux, ils pourront servir d'"ardoises électroniques" ou encore de tables de jeux pour partenaires éloignés, d'échiquiers télécommandés, par exemple.

Le service pourra faciliter et accélérer la lecture des journaux, la correspondance, la publicité, les opérations bancaires et toutes sortes de jeux électroniques. L'accès du public à l'information sera l'une des premières applications du service vidéotex étant donné qu'un grand nombre de nos activités sont reliées à l'intérêt que nous portons aux événements qui se produisent autour de nous. De plus, la technologie nécessaire à cette application est déjà bien comprise. Enfin, parce qu'il peut répondre à un grand nombre de besoins particuliers, un service de recherche documentaire possède le potentiel requis pour intéresser un marché suffisamment vaste.

Les promoteurs du vidéotex préconisent une technique de recherche par arborescences pour accéder à l'information. Ce sont les Postes britanniques qui ont tout d'abord proposé et mis en oeuvre cette technique (Fédida, 1975). Son principal avantage est la faible consommation qu'elle fait des ressources informatiques pour répondre à un besoin d'information. Ce facteur est important lorsque des milliers d'utilisateurs sont reliés à l'ordinateur en même temps. Plus l'ordinateur doit accomplir de fonctions pour répondre à une demande, plus l'utilisateur doit attendre longtemps avant d'obtenir la page demandée et moins le système lui semble interactif. Comme exemple d'une recherche dans une structure arborescente, lorsqu'un utilisateur choisirait l'article "Immeuble" dans une des pages de menu, celle-ci serait immédiatement suivie d'une autre

liste contenant les noms des secteurs de la ville. Le choix qu'il ferait de n'importe quel des articles de cette dernière page déclencherait l'apparition d'une autre page affichant des catégories d'immeubles résidentiels et commerciaux. Par une série de choix successifs dans cette page et dans des menus subséquents, l'utilisateur pourrait finalement obtenir la description d'une maison sise à une adresse particulière. L'information pourra avoir été introduite dans le système soit par une société immobilière, soit par une personne désirant vendre sa propre maison. Dans un cas comme dans l'autre, toute personne suffisamment intéressée par la description trouvée de façon si commode pourrait communiquer ensuite directement avec l'auteur de l'information.

La question de la description de l'image

Plusieurs pays comme la Grande-Bretagne, la France et le Canada ont orienté leur démarche dans deux sens différents relativement aux systèmes vidéotex. La principale différence repose sur la façon de coder les images aux fins d'entreposage dans la base de données et de transmission ultérieure aux écrans terminaux. Les systèmes européens sont des systèmes à adressage direct aux caractères qui limitent l'affichage à des textes à structure fixe et à des images graphiques rudimentaires. Si cette démarche prévaut, il sera difficile, voire impossible, de modifier les méthodes de transmission pour permettre l'affichage d'images de qualité supérieure.

La capacité qu'ont les systèmes à adressage direct aux caractères de s'adapter aux nouvelles exigences est restreinte surtout par la façon dont l'image est décrite. Les images sont transmises sous forme de portions séquentielles d'une image qui comprend 24 lignes de 40 caractères. Les images graphiques sont construites de la même façon à partir de caractères graphiques

codés qui sont assemblés pour former une mosaïque d'éléments visuels. Cette technique exige que l'information stockée comprenne des données sur le degré de résolution du terminal d'affichage. Par conséquent, les banques de données seraient tenues de stocker diverses versions d'une image afin de pouvoir la transmettre aux terminaux à résolutions différentes.

Pour qu'un système vidéotex soit à la hauteur de la technologie d'aujourd'hui, il doit être conçu de façon à pouvoir incorporer facilement les perfectionnements futurs. Cette compatibilité exige que les images décrites dans la base de données soient entièrement indépendantes de la méthode d'accès aux données, des caractéristiques du moyen de transmission et de la construction du terminal d'affichage. Les instructions de description de l'image (IDI) mises au point par le Centre de recherches sur les communications du ministère des Communications du Canada et incorporées dans le système vidéotex Télidon (Bown, O'Brien, Sawchuk et Storey, 1978) répondent à ces critères. Les IDI décrivent les images graphiques d'après les formes géométriques qu'elles contiennent. Il suffit de sept instructions de base, suivies de données très précises pour décrire la plupart des images graphiques. Quatre de ces instructions définissent les objets en termes de figures géométriques primaires qui comprennent des lignes, des arcs de cercle, des surfaces rectangulaires ou des polygones. Une cinquième instruction indique que les données sont transmises sous forme de bits, ou en mode photographique, dans le cas des images ne pouvant être décrites au moyen des quatre figures géométriques primaires. Une sixième instruction définit la position de l'objet à l'écran. La septième instruction sert au contrôle. Elle établit généralement un registre d'état avant la transmission d'autres instructions. Par exemple, l'une des fonctions de l'instruction de contrôle consiste à définir la couleur d'un objet. Les sept instructions peuvent définir les images sous une forme compacte aux fins de transmission à un terminal conçu pour les interpréter.

Un autre ordre est nécessaire si l'on veut passer du mode graphique au mode alphanumérique pour la transmission de textes et si on veut revenir au mode graphique pour la transmission d'information sous forme graphique. Le mode alphanumérique est le mode de fonctionnement normal, de sorte qu'un sous-ensemble du jeu d'IDI peut être utilisé avec des terminaux simples qui ne répondent intelligiblement qu'aux informations alphanumériques.

Les codes IDI proposés ne seront utiles que si le terminal récepteur possède l'intelligence voulue pour les interpréter et remplir la mémoire de visualisation des contenus appropriés. C'est pourquoi le terminal doit être muni d'un microprocesseur pour interpréter les instructions et produire le code de mémoire de visualisation. Quant à la mémoire de visualisation elle-même, il peut s'agir d'une mémoire avec adressage direct aux caractères ou d'une mémoire travaillant au niveau du bit. En Europe, on a conçu un terminal avec adressage direct aux caractères. Le terminal du système Télidon canadien, pour sa part, utilise une mémoire de visualisation par profil de multiplets. Dans ce cas, chaque élément imagier présenté à l'écran correspond à une position précise en mémoire. La couleur et la brillance de l'élément imagier dépendent du contenu de la position à laquelle il correspond. Cette démarche exige plus d'éléments de mémorisation que celle de la mémoire avec adressage direct aux caractères, ce qui en augmente le coût. Il faut néanmoins comparer le coût accru aux avantages que comporte une plus grande précision de l'image. De plus, le coût des mémoires ne cesse de diminuer grâce à l'amélioration des techniques de circuits intégrés. La figure 1 montre la qualité graphique obtenue au moyen du système Télidon à résolution d'affichage de 240 éléments imagiers sur 320.

Placer la figure 1 à peu près ici

Le code IDI sur lequel est fondé le système Télidon est un code souple. Dès que l'on concevra de nouveaux usages pour les terminaux vidéotex, le code pourra être élargi de façon à incorporer de nouvelles instructions qui faciliteront des services comme la production d'images et la communication de textes et d'images graphiques en mode interactif entre abonnés.

On envisage déjà l'addition d'instructions de traitement de l'image (ITI) à l'usage des fournisseurs d'information. Ces instructions permettraient à l'utilisateur d'un terminal semblable au terminal Télidon de produire des images sur son écran de télévision et d'en stocker les descripteurs dans un fichier de données. Il serait possible d'effectuer l'effacement sélectif d'images produites antérieurement ainsi que d'y apporter des additions et des modifications. D'autres instructions encore permettraient de faire tourner sur elles-mêmes, de reproduire à l'échelle ou de transposer des portions de l'image visualisée. Il se peut même que des instructions soient mises au point pour coder les structures phonétiques de base afin de synthétiser la parole pour accompagner les images visualisées.

Le Centre de recherches sur les communications poursuit ses recherches sur la transmission des images entre deux personnes (Bown, O'Brien, Warburton et Thorgeirson, 1975). Cette application serait utile à des interlocuteurs qui souhaiteraient discuter les mérites d'une structure spatiale quelconque. Les échanges entre un architecte et un client au sujet des plans d'un bâtiment en sont un exemple. D'autres applications pourraient se présenter dans le domaine

des divertissements, où des jeux pourraient être adaptés ou inventés en vue de tirer parti de la capacité de communiquer en mode interactif. Tant que la technique de description de l'image est indépendante du matériel utilisé, les perfectionnements technologiques ne sauraient condamner une base de données à la désuétude.

Questions en rapport avec l'utilisateur

Le perfectionnement technique d'un produit de consommation oblige les concepteurs à s'interroger sur ce qui conviendra le mieux aux utilisateurs. Avant la mise en oeuvre de services vidéotex, non seulement doit-on déterminer quels scnt ceux qui conviennent le mieux au moyen, mais on doit aussi s'arrêter aux questions que soulèvent tant l'affichage de l'information que le protocole de transmission interactive.

L'affichage de l'information soulève des questions parce qu'on utilisera le téléviseur ordinaire pour présenter des images fixes au lieu des images dynamiques pour lesquelles il a été conçu. La plupart des données que nous possédons sur les facteurs humains en rapport avec les questions d'affichage ne peuvent être appliquées directement aux écrans de télévision à matrice. Ou bien on a employé la technologie de la télévision analogique ou bien on n'a employé aucun écran électronique. Il n'y a aucun doute que les écrans de télévision numérique possèdent des caractéristiques nouvelles et différentes qui doivent être étudiées en propre. De plus, la nouvelle technologie se prête à des expériences difficiles à faire auparavant. Elle permet aussi l'exploitation économique de certaines caractéristiques, comme la couleur. En raison du coût prohibitif de l'emploi de la couleur dans le traitement de l'information imprimée, on n'a pas toujours pu étudier à fond le potentiel du codage chromatique.

Des négociations sont en cours entre un certain nombre de pays en vue de normaliser, notamment, la façon de composer les textes à l'écran; de telles normes sont appelées à influer sur l'acceptabilité du support d'information. Toute norme que propose un pays, quel qu'il soit, doit s'appuyer sur des données empiriques probantes. C'est pourquoi nous étudions présentement dans quelle mesure divers paramètres d'affichage reliés à la densité des caractères sur une page altèrent la lisibilité d'un texte. Nos expériences constitueront la base d'une proposition portant sur le nombre maximal de caractères par ligne et sur le nombre maximal de lignes par page convenant le mieux à la télévision nord-américaine. Une telle norme ne peut cependant s'appliquer que dans des conditions bien précises. Par exemple, la densité optimale de caractères peut varier selon qu'il s'agit d'un écran de télévision monochrome ou polychrone, d'une entrée RF ou RVN, ou d'un tube-image à matrice à pointillé ou linéaire. Elle peut différer selon qu'on emploie un espacement proportionnel des caractères ou non. D'autres facteurs environnementaux, notamment l'éclairage ambiant, la distance d'observation, le contraste de l'image (tant la luminance que le contraste des couleurs) et la dimension de l'écran peuvent aussi déterminer la densité optimale de caractères au regard de la lisibilité d'un texte. Nous avons aussi des preuves que la mesure dépendante peut modifier du tout au tout le résultat d'une expérience portant sur les caractéristiques d'affichage. Plus précisément, il a été démontré que la vitesse de lecture augmente lorsque les caractères sont plus petits, alors que le temps de recherche d'un caractère-cible placé au hasard dans un espace à deux dimensions diminue lorsque les caractères sont plus gros (Snyder et Maddox, 1978). Il se pourrait donc que la lisibilité optimale exige des caractères de grosseurs différentes pour les textes suivis et pour les annotations graphiques. Ce qui semblait de prime abord être une

simple question de nombre optimal de caractères par ligne et de nombre optimal de lignes par page est devenu beaucoup plus complexe en raison des divers autres facteurs qui doivent être pris en considération.

D'autres questions relatives à l'affichage ne relèvent pas directement de l'adoption de normes internationales pour le vidéotex. Elles sont néanmoins importantes pour l'acceptation du support d'information par l'utilisateur. Ainsi, l'examen des mérites relatifs des écrans à faible et à grande résolution devrait tenir compte de l'écart de prix entre les deux options. Il pourrait fort bien exister un point au-delà duquel, pour la plupart des applications, on ne gagne rien à augmenter la finesse de résolution. En outre, il y a eu très peu de recherches sur l'optimisation d'un jeu de caractères de bas de casse pour les écrans à matrice à pointillé. Il y aurait lieu de concevoir un certain nombre de polices de caractères de bas et de haut de casse de taille et de forme différentes, et de les évaluer en vue de leur utilisation par les fournisseurs d'information. Il importe d'établir des principes régissant la conception graphique de façon à ce que les fournisseurs d'information puissent créer des pages d'information claires et agréables. Un emploi judicieux des couleurs et une bonne organisation spatiale des pages peuvent en faciliter la lecture. Il est nécessaire que des règles de clôture psychologique soient formulées explicitement pour éviter qu'un concept incomplet demeure en suspens lorsqu'il doit s'écouler entre cinq et dix secondes avant la réception d'une nouvelle page. On doit aussi examiner l'utilité comparée des diverses façons dont un texte peut se dérouler en continu sur l'écran.

Le protocole qui régira le dialogue pour fins de recherche documentaire est une autre question complexe, étant donné qu'il touche la structure de la base

de données, le langage et le mode d'interrogation, les périphériques d'entrée et l'organisation de l'information dans l'esprit des gens. L'extraction des données stockées suivant une structure hiérarchique ou arborescente est techniquement plus efficace parce que c'est la structure dont la consultation consomme le moins de ressources informatiques. Cette considération est importante car une réponse trop lente de l'ordinateur principal peut éliminer l'aspect interactif qui devrait caractériser les services vidéotex. On peut toutefois s'interroger sur l'utilité d'une réponse rapide dont le contenu informationnel serait minime. La consultation successive d'une longue série de pages de menu peut ressembler à une conversation avec une personne à la réplique rapide mais au vocabulaire très restreint. Il devient alors laborieux d'extirper le message désiré. La structure arborescente est efficace sur le plan technique, mais elle peut être très inefficace du point de vue de l'utilisateur qui veut obtenir un renseignement ayant de multiples attributs. Les recherches dans ce domaine devraient s'appuyer sur l'une ou l'autre des hypothèses suivantes: (1) du point de vue technique, il est impossible d'envisager d'autres méthodes que la structure hiérarchique, et (2) des structures de réseau plus raffinées sont possibles.

Suivant la première hypothèse, les chercheurs doivent tenter de déterminer des règles d'organisation hiérarchique de l'information de sorte qu'il y ait le moins d'ambiguité possible dans l'esprit des gens quant au cheminement susceptible d'aboutir au renseignement voulu. Ceci revient à se demander s'il y a meilleure congruence de la carte mentale de différentes personnes avec certaines structures hiérarchiques qu'avec d'autres. Suivant la seconde hypothèse, un effort interdisciplinaire s'impose pour définir les caractéristiques d'un système conçu pour répondre le plus rapidement possible

à des demandes de recherches documentaires. Dans ce cas comme dans le cas précédent, on doit étudier de façon générale la carte cognitive des gens. Des connaissances de ce genre peuvent aider d'autres spécialistes, notamment les analystes fonctionnels, à concevoir les moyens de recherche documentaire qui sont les plus efficaces du point de vue de l'utilisateur.

Le langage et le mode (c.-à-d. sélectif contre génératif) d'interrogation et la conception du périphérique d'entrée le plus approprié sont d'autres questions pertinentes et interdépendantes, sur lesquelles la structure logique de l'information dans la base de données exerce en outre une influence marquée. On peut présumer qu'une grande partie des utilisateurs possibles ignorent tout des ordinateurs et, en fait, ne souhaitent pas apprendre une procédure compliquée d'accès à l'information. Il incombe dès lors aux analystes organiques, et peut-être aux fournisseurs d'information, de simplifier l'accès à l'information tout en diminuant le temps d'attente de l'information recherchée.

Ce qui déterminera peut-être, finalement, la viabilité des services vidéotex dans une économie de marché libre, c'est la façon dont le support d'information sera employé. Il possède plusieurs caractéristiques qui en font un système relativement unique en son genre. Il est interactif, axé sur le réseau, et se prête facilement à l'emploi de la couleur. Il importe d'exploiter ces caractéristiques de façon que le vidéotex puisse être perçu comme un support offrant à des populations particulières d'utilisateurs des services à la fois utiles et uniques.

Documentation

Bown, H.G., O'Brien, C.D., Sawchuk, W. et Storey, J.R. Description générale du Télidon: proposition canadienne concernant les systèmes vidéotex. Note technique n° 697-F du CRC, ministère des Communications, Canada, 1978.

Bown, H.G., O'Brien, C.D., Warburton, R.E. et Thorgeison, G.W., System independence for interactive computer graphics application programs. Proceedings of the 4th Man-Computer Communications Conference, Ottawa, May 26 and 27, 1975.

Fedida, S., Viewdata: an interactive information service for the general public. Proceedings of the European Computing Conference on Communications Network, 1975.

Snyder, H.L. et Maddox, M.E., Information transfer from computer-generated dot-matric displays. Report HFL-78-3/ARO-78-1, U.S. Army Research Office, Research Triangle Park, N.C., 1978.

Figure 1

Exemple de la qualité des graphiques que permet le système vidéotex Télidon.

Herbert G. Bown est directeur de Systèmes et réseaux de données au Centre de recherches sur les communications (CRC) du ministère des Communications. Au cours des 12 années qu'il a passées au CRC, il a participé à de nombreux projets dans le domaine des télécommunications. Il possède une maîtrise ès sciences en génie électrique du Collège technique de la Nouvelle-Écosse.

C. Douglas O'Brien est un ingénieur de recherche affecté au projet du Télidon. Il a obtenu sa maîtrise en génie des systèmes à l'Université Carleton et travaille au CRC depuis trois ans.

William Sawchuk est directeur de programme de la transmission des images au CRC, où il travaille depuis six ans. Il possède une maîtrise ès sciences de l'Université de la Saskatchewan et un Ph.D. en physique, de l'Université de Calgary.

John R. Storey est le chef des recherches sur les terminaux au CRC. Il travaille depuis de nombreuses années dans les domaines de l'électronique et des télécommunications.

William C. Treurniet est un psychologue de recherche du CRC. Avant d'être affecté aux recherches sur les facteurs humains rattachés aux systèmes vidéotex, il a aidé à évaluer les essais sur le terrain des systèmes de transmission de la voix humaine au moyen du satellite Hermès. Il possède un baccalauréat ès sciences en physique et une maîtrise ès arts en psychologie de l'Université de Waterloo.

Pour communiquer avec l'un ou l'autre des auteurs, s'adresser à:

Centre de recherches sur les communications
Ministère des Communications
C.P. 11490, succursale H
Ottawa (Ontario), Canada
K2H 8S2

Répercussions de Télidon dans le monde de l'information

par Dorothy Phillips

Ministère des Communications

Ottawa

Septembre 1979

Présenté devant l'Association canadienne des sciences de l'information (section de Montréal), le 25 septembre 1979.

Aussi disponible en anglais

Je suis heureuse d'assister à votre réunion de ce soir et de vous parler de quelques-unes des conséquences qu'amènera l'avénement de Télidon et d'autres réalisations semblables dans le monde de l'information. Il est indéniable que Télidon a fait surgir une foule de questions quant à la manière de concevoir les systèmes électroniques d'information afin de les rendre le plus acceptables possible pour leurs usagers. Or, au rythme où la technologie évolue, il faudra prendre en la matière des décisions qui toucheront les usagers et l'économie, sans se donner le temps de faire toutes les recherches nécessaires. Etant donné qu'un bon nombre de ces questions visent la façon de concevoir et de répertorier l'information pour en faciliter l'accès aux usagers, préoccupations qui font partie de votre travail de tous les jours, j'aimerais ce soir vous communiquer un peu de l'enthousiasme qui nous anime pour les recherches dans ce domaine et solliciter votre aide pour que nous y trouvions des réponses assez tôt pour influencer dans une certaine mesure les progrès techniques des prochaines années.

Modèle relatif aux questions soulevées par les recherches sur Télidon

Avant d'aborder des questions précises sur la conception des banques de données à l'intention des usagers de Télidon, permettez-moi de vous présenter un modèle général relatif aux questions soulevées par les recherches sur Télidon, pour que vous ayez une bonne idée de leur complexité. (Figure 1). Ce modèle indique les différents éléments dont il faut tenir compte dans la mise au point de l'interaction du système Télidon et dans quelle mesure chacun doit tenir compte du point de vue des autres au moment de prendre des décisions. En outre, il sera peut-être bon de nous rappeler que les questions à résoudre sont trop nombreuses et que le choix judicieux des questions à aborder les premières peut être aussi important que l'emploi des bonnes méthodes.

MODÈLE RELATIF AUX QUESTIONS SOULEVÉES PAR LES RECHERCHES SUR TÉLÉDON

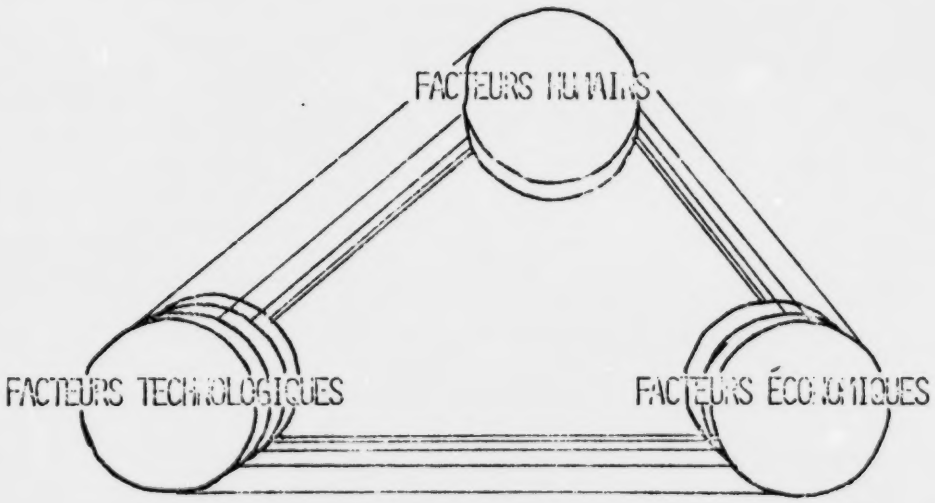


FIGURE 1

Télidon comporte des aspects humains, économiques et techniques. On ne peut prendre une décision sur l'un de ces aspects sans tenir compte des répercussions qu'elle aura sur les deux autres. En effet, chacun de ces aspects comporte des possibilités et des limites qui lui sont propres. Par exemple, les progrès techniques ont fait figure de proue en multipliant chaque année les possibilités de réalisations, mais ils sont eux mêmes conditionnés tant par l'économie que par les possibilités et les limites de l'homme.

La mise au point du service vidéotex doit susciter l'intérêt des usagers de façon à devenir économiquement rentable pour permettre à ses réalisateurs de recouvrer les frais de mise en oeuvre qu'il aura coûtés et de songer à mettre en oeuvre d'autres nouveautés techniques. Quant à la question de savoir quelles nouveautés techniques il importe de mettre en oeuvre en premier lieu, on se fondera sur les préférences des usagers et sur les possibilités d'action permises par la conjoncture économique.

Le diagramme de la Figure 1 décrit l'interaction de ces trois aspects de la mise au point de Télidon et l'évolution rapide des aspects économiques et techniques. De bien des points de vue, l'homme change relativement peu pendant de courtes périodes de temps; ses mécanismes de perception et de compréhension sont assez stables. Étant donné que ses aspirations dans certains domaines culturels pourront changer au fur et à mesure de son apprentissage, le modèle tient compte jusqu'à un certain point de ces changements qui peuvent s'opérer en lui.

Lorsque nous cherchons à déterminer quelles questions étudier en détail, il est important de nous rappeler cet aspect de l'évolution dans les domaines connexes. Il est peut-être intéressant et important de connaître les réactions de l'homme vis-à-vis d'une réalisation technique particulière, mais si une autre réalisation vient la supplanter l'année suivante, la recherche peut devenir désuette à très brève échéance, ce qui signifie que nous aurons perdu notre temps. Nous cherchons à faire en sorte que la recherche soit générale et ne se limite pas à un dispositif particulier, mais ce n'est pas toujours possible. Il faut se tenir au fait des perspectives d'évolution de la technologie.

La plupart des personnes qui participent à la mise au point de Télidon ont une formation de technicien, d'économiste ou de chercheur en sciences sociales, mais comme chacune des questions à l'étude touche au moins deux aspects du modèle, cet outil nous permettra d'étudier les questions sous deux aspects à la fois et d'ignorer le troisième. Cette façon de procéder semble nécessaire en raison de la complexité des problèmes.

En tant que psychologues, nous nous préoccupons surtout de l'interaction entre l'homme et la technologie. Nous savons bien que notre travail a des répercussions économiques, mais nous en faisons abstraction afin de circonscrire les questions et d'en faciliter l'étude. D'autres étudient l'interaction entre la technologie et l'économie ou celle entre l'homme et l'économie, tout en ignorant temporairement le troisième pôle du triangle.

Je crois que la plupart des travaux des scientifiques du monde de l'information se sont donné pour objet l'interaction entre l'homme et la technologie ou l'interaction entre la technologie et l'économie. Permettez-moi de m'arrêter maintenant aux questions que Télidon a soulevées au sujet de l'interaction entre l'homme et la technologie, car un bon nombre d'entre elles intéressent directement les scientifiques du monde de l'information.

Questions soulevées par Télidon sur l'interaction entre l'homme et la technologie

L'interaction entre l'homme et la technologie des communications peut être envisagée sous plusieurs aspects. D'une part, l'aptitude de l'homme à voir et à entendre impose des limites assez sérieuses à la conception technique, mais comme ces limites sont assez bien connues, nous n'avons plus tellement de questions à nous poser à ce sujet. D'autre part, l'aptitude de l'homme à recevoir et à traiter l'information et à y répondre par des gestes ou des paroles impose également des limites et c'est ici que se situent bon nombre de nos questions. Par ailleurs, les aspirations et les besoins culturels et sociaux de l'homme proposent différentes orientations en vue du perfectionnement des communications du point de vue technique. Dans une bonne mesure, les questions que nous nous posons à ce stade particulier du perfectionnement de Télidon demeurent encore imprécises et nous devrons les préciser davantage avant d'entreprendre des recherches empiriques.

Notre réflexion s'est attachée jusqu'à maintenant au traitement de l'information du point de vue humain et, dans une certaine mesure, aux aspérations sociales et culturelles de l'homme. Cependant, il y a trois questions de base dont j'aimerais vous entretenir à propos de la mise au oeuvre de Télidon, pour nous situer par rapport à elles. Ces questions sont les suivantes:

1. Quels renseignements faudraient-ils présenter à l'aide de Télidon à l'intention des utilisateurs privés?
2. Comment faudrait-il présenter une page d'information de Télidon?
3. Comment faudrait-il établir un dendrogramme ou index ramifié?

Déjà, vous vous rendez probablement compte qu'il ne s'agit pas là de questions entièrement distinctes. En effet, l'établissement d'un index ramifié est fonction des données contenues dans le fichier central et la conception d'une page est elle-même déterminée par la place qu'elle occupera dans l'ensemble.

1. Quels renseignements devraient être présentés à l'intention des usagers privés?

Jusqu'à maintenant, en créant une banque de données pour fins de démonstration et de recherche, nous avons surtout songé à recueillir des renseignements susceptibles d'intéresser un nombre assez considérable de personnes. Nous y avons donc introduit des spécimens de données de chaque catégorie d'information si bien que la banque de données comprend maintenant plusieurs catégories de renseignements: les nouvelles d'actualité, météorologiques et sportives, les services du gouvernement, des données historiques,

des renseignements à l'intention des consommateurs, des guides touristiques, gastronomiques et sur les spectacles, des guides du marché des valeurs boursières et des affaires, des offres d'emploi, des jeux, des bandes dessinées et des cours sur la respiration artificielle.

Une autre catégorie de renseignements à laquelle nous avons pensé, mais que nous n'avons pas encore créée, est la catégorie de l'information communautaire, qui comprendra dès renseignements sur les activités, les clubs, les organismes sociaux, les organismes d'éducation, les lois et les services gouvernementaux locaux, ainsi qu'une foule de renseignements supplémentaires.

Tous ces services font l'objet de la recherche documentaire. Vous connaissez bien les services électroniques de recherche documentaire, comme les banques de données bibliographiques. Dans quelle mesure pourrait-on y avoir accès par l'intermédiaire de Télidon? Faudrait-il établir un service personnalisé à l'intention d'un groupe restreint de clients? Pour répondre à cette question, j'aimerais me reporter au modèle dont je vous ai parlé précédemment, car un certain nombre de questions pourraient être posées concernant chacune des interactions du triangle. Pour ce qui est de l'interaction entre l'homme et la technologie, on peut se demander si les recherches de données bibliographiques peuvent se simplifier au point où l'usager profane puisse trouver les renseignements qu'il désire à la maison sans aide technique. Pour ce qui est de l'interaction entre l'homme et l'économie, les gens qui désireront bénéficier de ce service seront-ils suffisamment nombreux pour qu'il soit possible de l'offrir de façon économique? Enfin, en ce qui concerne

l'interaction entre la technologie et l'économie, peut-on réaliser les progrès techniques à un coût raisonnable? Il faudra donc effectuer une foule de recherches pour savoir si les services de recherche documentaire dans le domaine de la bibliographie seront viables.

Pour des raisons économiques et techniques, Télidon doit s'en tenir présentement à une banque de données à structure ramifiée, ce qui nous astreint à travailler avec des services de recherche documentaire à structure ramifiée. Toutefois, si l'on se reporte encore une fois au modèle, il est important de se rappeler que si l'on modifie l'interaction entre la technologie et l'économie, on pourra offrir des services plus perfectionnés. Le jour où d'autres structures seront adaptées à la recherche documentaire (comme les recherches à partir de mots-clés et les demandes de renseignements en langage courant), nous nous poserons encore les questions que nous nous posons aujourd'hui par rapport à la structure ramifiée et nous nous demanderons en outre si ces nouvelles techniques constituent un progrès par rapport à la structure actuelle. Il est fort probable que la réponse à la dernière question soit complexe, les autres techniques de recherche permettant d'assurer un meilleur service aux usagers pour certains genres d'information.

En outre, certains services qui ne peuvent pas être intégrés dans les services de recherche documentaire peuvent constituer des applications importantes de Télidon. Par exemple, l'enseignement automatisé se perfectionne depuis quelques années et il existe maintenant un quantité importante de logiciel didactique. Pour se prévaloir de ce

service, l'usager aura besoin d'un clavier complet et d'un ordinateur qui analysera sa réaction et donnera à ses questions des réponses de nature à accroître ses connaissances. La plupart des programmes d'enseignement automatisé n'ont pas eu recours aux graphiques et aucun graphique de ce genre n'a été préparé pour Télidon. Nous devrons bientôt trouver la réponse à plusieurs questions, par exemple, dans quelle mesure les gens seront-ils prêts à payer pour obtenir les programmes d'enseignement automatisé? De quelle façon pouvons-nous déterminer si les programmes conçus par Télidon intéresseront les usagers? Les progrès techniques visant à introduire l'enseignement automatisé dans les foyers au moyen de Télidon peuvent-ils se réaliser malgré les restrictions économiques?

Après quelque temps de réflexion sur l'évolution de Télidon et de l'enseignement automatisé, nous élaborons présentement les exigences techniques à respecter pour la création de cours complexes d'enseignement automatisé pour Télidon.

Nous pouvons faire une analyse semblable dans le cas des jeux interactifs, car ce service nécessitera des analyses informatisées. Ces jeux où l'usager relève le défi d'une machine dont il ne contrôle que quelques paramètres, semblent très répandus parmi les amateurs d'informatique. Ce service pourrait intéresser les usagers privés.

2. Comment devrait-on présenter les pages d'information de Télidon?

Le concepteur devra connaître en détail les limites et les capacités de Télidon. Présentement, une page affichée à l'écran peut contenir au maximum 40 caractères par ligne et au plus 20 lignes, soit 800 caractères, ce qui représente environ 120 mots. Cependant, pour soutenir l'intérêt de l'usager, le texte ne doit pas remplir la page en entier. En effet, l'agencement des graphiques et des couleurs et la disposition de la page sont des éléments importants pour créer une impression esthétique agréable et attirer tout de suite l'attention des usagers sur le renseignement important à retenir. Télidon offre 8 couleurs, y compris le blanc et le noir, et 6 tons de gris, mais nous avons constaté qu'une page de texte qui utilise plus de trois couleurs présente un aspect confus. En outre, le concepteur devra choisir la couleur de fond du texte et des graphiques. Les couleurs ayant différents degrés de luminance, il semble que les meilleurs combinaisons de couleurs de fond et de surface sont celles qui offrent le contraste le plus frappant.

Une bonne partie des notions que les typographes et les graphistes possèdent au sujet de la disposition des textes et des graphiques s'applique aux imprimés et ne convient pas directement aux publications électroniques. Dans le cas de Télidon, il faut mettre au point des lignes directrices sur l'utilisation de l'espace, des couleurs et des fontes. Nous préparons présentement un premier projet de lignes directrices, mais il nous faudra du temps pour acquérir la confiance des experts dans le domaine.

Au cours des recherches que nous avons menées jusqu'à maintenant, nous nous sommes attachés à des problèmes particuliers, comme les dimensions des caractères et leur espacement minimal pour en assurer la lisibilité. Nous avons jusqu'à maintenant recueilli des données sur l'espacement idéal de deux fontes. Les paramètres d'espacement feront partie du système de façon permanente, ce qui devrait éliminer un souci pour les concepteurs.

3. Comment devrait-on établir un dendrogramme ou index ramifié?

Les limites et les possibilités de la structure ramifiée de Télidon doivent entrer en ligne de compte lorsqu'il s'agit de diffuser l'information. La structure actuelle de neuf niveaux offre 9 options par page. Une table de renvois permet d'avoir accès aux pages soit directement, soit au moyen de la structure ramifiée. Pour cette raison, le concepteur aura plusieurs décisions à prendre. Il devra décider du nombre d'options à utiliser à chaque page et du degré de spécialisation de l'information. Il pourra également ranger l'information dans des documents qui comptent jusqu'à 1 000 pages d'information séquentielle. Le ministère des Communications a entrepris une analyse destinée à déterminer le nombre optimal d'options à offrir dans une page. Si l'on tient compte du temps qu'il faut consacrer à la lecture des options et du temps d'accès aux pages de l'index, le modèle établi permet de calculer le nombre d'options sur une page, ce qui réduira à un minimum le temps total à consacrer à la recherche. D'après ce modèle, le nombre optimal d'options utilisées sur une page varie entre trois et six, et il semble que l'utilisation de plus de dix options augmente le temps de recherche de façon appréciable. Nous n'avons pas encore mis le modèle à l'essai.

Nous devrons également décider quelles catégories nous établirons et de quels indicateurs nous nous servirons pour classer notre information dans une structure. Dans bien des cas, la question est résolue à l'aide d'un index. Nous en avons d'ailleurs étudié plusieurs pour déterminer dans quelle mesure ils pouvaient nous servir, mais, dans la plupart des cas, des index ne permettent que deux ou trois degrés de spécialisation et, de toute façon, l'information traitée ne semble pas convenir au matériel que nous avons décidé d'introduire dans la banque de données créée pour fins de démonstration. Nous avons également étudié la constitution de ces index pour nous en inspirer, mais il semble que la plupart d'entre eux aient été établis à partir de listes antérieures de termes utilisés ou à partir d'un ensemble bien connu de renseignements appartenant à une série assez constante de catégories, comme un journal.

Lorsque nous avons créé la banque de données de Télidon, nous n'avions aucun ensemble de renseignements structuré de la sorte. La banque de données s'est accrue au fur et à mesure que nous avons trouvé des renseignements à y inclure et, au besoin, nous avons ajouté des catégories supplémentaires aux niveaux supérieurs de la structure ramifiée et nous en avons même créé de nouvelles à des niveaux inférieurs afin de classer les renseignements dans les catégories correspondantes, comme si nous avions dessiné un arbre en commençant par le tronc et en allant à la fois vers le feuillage et vers les racines, ce qui ne produit pas nécessairement une structure bien ordonnée. Nous avons maintenant entrepris l'étude de la structure ramifiée de la banque de données créée pour fins de démonstration, pour déterminer les endroits où les usagers font le plus d'erreurs dans les recherche de renseignements. En étudiant

des questions de peu d'envergure auxquelles nous pouvons répondre rapidement, nous espérons déterminer si la structure ramifiée fonctionne bien et apprendre à établir des notions générales sur la façon de mettre au point une bonne structure ramifiée.

Au cours de ces études, nous avons examiné quelques-unes des mesures relatives aux systèmes de recherche documentaire que nous retrouvons dans les documents sur les sciences de l'information. Les mesures de rappel[1] et de pertinence[2] et le facteur "nouvelles connaissances"[3] n'ont pas été utiles, parce qu'elles supposent que l'usager est à la recherche d'une série de documents. Dans le cas de Télidon, nous avons supposé que l'usager cherche un document unique. Nous avons donc utilisé certains facteurs, comme le temps nécessaire pour obtenir le matériel pertinent, le nombre d'erreurs commises et l'attitude des usagers. Si la banque de données était plus vaste, on pourrait supposer que l'usager cherche une série de documents et non pas un seul. La question des mesures pertinentes destinées à évaluer la banque des données de Télidon est donc loin d'être réglée.

1 mesure de rappel proportion de documents pertinents d'une base de données rappelés au cours d'une recherche.

2 mesure de pertinence proportion de documents rappelés au cours d'une recherche et jugés pertinents.

3 facteur "nouvelles connaissances" proportion des renseignements rappelés que l'usager juge pertinents et qui représentent pour lui de nouvelles connaissances.

Une partie des travaux effectués pour comprendre comment créer une banque de données de Télidon à structure ramifiée vise à déterminer la possibilité d'établir une structure de base normalisée pour toutes les banques de données, ce qui permettrait aux usagers à travers le pays de savoir où trouver l'information au cas où ils déménageraient d'une région à l'autre du pays. Si cette idée semble bonne à première vue, nous n'avons pas encore bien compris ce qu'il faut faire pour établir cette structure ou en connaître les éléments constitutifs. Cette structure devrait peut-être comprendre un ou plusieurs index, ou elle pourrait encore comprendre un ensemble de lignes directrices sur la façon d'organizer l'information.

Il se peut que tout le mal que nous nous donnons pour préparer des banques de données à structure ramifiée soit bientôt relégué aux oubliettes. Si les progrès techniques se font assez rapidement pour que nous puissions nous servir d'une structure de réseau, d'index établis à partir de mots-clés ou d'une structure de langage courant, ces efforts pourraient devenir désuets. Toutefois, une fois que ces progrès techniques seront réalisables, nous devrons nous demander si les usagers désirent toujours une banque de données à structure ramifiée ou si de nouvelles méthodes répondraient mieux à leurs besoins. La réponse à cette question dépendra peut-être du genre d'information que contiendra la banque de données. Et ceci nous ramène au problème que nous avons posé au départ, c'est-à-dire quelle information faut-il présenter au moyen de Télidon?

En résumé, dans chaque domaine du perfectionnement de Télidon, les facteurs humains, techniques et économiques réagissent entre eux de façon complexe. Les travaux sur l'interaction entre l'homme et la technologie ont surtout porté sur la corrélation entre la technologie et les capacités de l'homme dans le domaine du traitement de l'information. Les principales questions qu'il faudra résoudre sans trop tarder sont les suivantes: quelle information Télidon devrait-il présenter?, comment devrait-on concevoir les pages d'information? et comment peut-on établir une banque de données à structure hiérarchique, alors que nous ne savons pas au départ quels renseignements y seront introduits?.

D'ici quelques années, il deviendra important de savoir si les usagers veulent utiliser d'autres structures de recherche de l'information que la structure ramifiée et si d'autres services, comme l'enseignement automatisé et les banques de données bibliographiques, constitueraient des applications utiles de Télidon.

Il s'agit là de questions complexes qui sont importantes pour l'avenir de Télidon et celui de l'industrie électronique de l'information. J'espère que vous serez prêts à nous aider à y trouver des réponses.

CHAPITRE II

FACTEURS HUMAINS RATTACHES A L'AFFICHAGE VISUEL

Espacement des caractères sur un écran de télévision (Jeu de caractères 5 sur 7)

William C. Treurniet

Centre de recherches sur les communications

C.P. 11490, Succursale "H"

Ottawa (Ontario), Canada

K2H 8S2

Prétirage

Exposé présenté à la Conference on Processing of Visible Language (Conférence sur le traitement du langage visuel) tenue à Niagara-on-the-Lake, le 4 septembre 1979, et qui sera reproduit dans les actes de la conférence.

Résumé

La densité d'un texte présenté sur un écran de télévision peut en altérer la lisibilité. Une expérience comprenant une épreuve de recherche et un jeu de caractères 5 sur 7 a permis d'établir les valeurs minimales d'espacement entre les caractères qui seraient nécessaires pour assurer la bonne lisibilité d'un texte. Nous fondant sur les résultats de notre expérience, nous recommandons d'utiliser un espace horizontal de deux éléments imagiers et un espace vertical de trois éléments imagiers. De plus, avec le jeu de caractères 5 sur 7, les caractères descendants (g, j, p, q et y) devraient dépasser le bas de la ligne d'au moins un élément imagier. Ces données indiquent que l'on ne devrait pas afficher des pages de plus de 20 lignes, à environ 40 caractères la ligne, sur les téléviseurs répondant aux normes NTSC.

Un certain nombre de pays, dont le Canada, présentent à l'heure actuelle au public un type de service vidéotex permettant à l'usager, qu'il soit dans son foyer ou au travail, de capter sur un écran de télévision à trame des données provenant d'une base de données. Certaines de ces données peuvent être présentées sous forme de graphiques, mais la plupart des données seront probablement sous forme de textes à lire. Les communications traversent souvent les frontières, d'où la nécessité de s'entendre sur une norme commune d'affichage des pages de texte; c'est-à-dire que lorsque dans un pays, on crée une page de texte, cette page doit être affichée de la même façon dans un autre pays. Aussi, l'établissement d'une norme internationale réglerait les questions de l'espacement horizontal des caractères, de l'espacement entre les lignes et du dépassement des caractères descendants. A cette fin, le présent document suggère les valeurs appropriées, d'après les mesures du comportement obtenues en laboratoire.

Etant donné que les usagers du système vont lire ce qui sera affiché sur leur écran de télévision, la mesure la plus adéquate semble, à prime abord, celle de la vitesse de lecture ou de compréhension. Tinker (1965) passe en revue un certain nombre d'études qui prenaient la vitesse de lecture comme mesure dépendante pour évaluer diverses caractéristiques de la page imprimée. Lors de ses expériences, il recourait ordinairement à un grand nombre de sujets pour comparer des caractéristiques comme la fonte des caractères, la couleur de la page et l'éclairage ambiant. Le test de Tinker sur la vitesse de lecture (1955) avait des exigences minimes de compréhension; aussi, on pensait qu'il était sensible aux changements au niveau de la perception. Plus récemment, Snyder et Maddox (1978) se sont servis d'une forme modifiée du même test afin de démontrer l'incidence sur la capacité de lecture des variations survenant

dans l'affichage des caractères à matrice de points sur un écran cathodique. En fait, ils ont démontré que plus l'élément de la matrice était gros, plus il augmentait le temps de lecture, que la forme carrée d'un élément réduisait le temps de lecture par rapport aux formes allongées, et qu'une augmentation du rapport entre l'espace et la grosseur de l'élément augmentait le temps de lecture. Cependant, des études menées dans notre propre laboratoire ont démontré que le test sur la vitesse de lecture était insensible au fait de présenter un texte en haut de casse ou en bas de casse soit sur papier, soit sur un écran de télévision. Auparavant, des études comparant l'effet de l'emploi de bas de casse et de haut de casse sur la vitesse de lecture avaient montré qu'un texte en bas de casse se lit plus rapidement qu'en haut de casse à des distances de lecture normales (Breland et Breland, 1944; Paterson et Tinker, 1946; Warren, 1942). Le fait que nous n'ayons pu prouver cet effet met alors en doute la possibilité qu'il se répète sur un écran cathodique et fait aussi douter de la fiabilité du test sur la vitesse de lecture. Au moins quelques-unes des premières recherches, en plus de l'étude de Snyder et Maddox (1978), étaient conçues sur l'interrelation entre les sujets, alors que notre recherche était centrée sur le sujet. Les résultats que nous avons obtenus peuvent donc avoir subi l'influence des effets reportés. Si le test sur la vitesse de lecture n'est fiable qu'avec l'interrelation entre les sujets, il ne peut constituer un outil efficace en ce qui concerne le nombre de sujets requis; il faudrait étudier la possibilité d'utiliser d'autres méthodes quand on doit tester un grand nombre de conditions expérimentales.

Une autre méthode utilisée pour évaluer la qualité de l'affichage repose sur une épreuve de recherche. Dans un tel cas, les mesures dépendantes sont, d'une part, le temps requis pour repérer un objectif précis après la présentation

de l'image et, d'autre part, la fréquence des erreurs dans l'identification de l'objectif. La valeur accordée au temps de recherche comme mesure de la qualité d'affichage pour la lecture de textes constitue une question méritant un examen approfondi. Snyder et Maddox (1978) ont démontré que des modifications apportées à leurs variables dépendantes ayant une incidence sur la grosseur des caractères entraînent des effets contraires sur le temps de recherche et de lecture du menu. On identifie plus rapidement les caractères plus gros que les caractères plus petits lorsque le contexte apporte peu d'aide; par contre, on lit moins rapidement les gros caractères lorsqu'ils s'accompagnent d'un contexte sémantique. D'après Snyder et Maddox (1978), les caractères plus gros augmentent le temps de lecture parce que l'oeil doit fixer davantage pour couvrir une étendue plus grande; en revanche, le temps de recherche diminue quand les caractères sont plus gros parce qu'ils sont plus faciles à repérer dans le champ visuel. Les conclusions de ces deux auteurs mettent en doute l'opportunité d'utiliser une épreuve de recherche de menu comme eux-mêmes l'ont fait, lorsque le rendement visé va de la perception générale à la lecture.

Bien que la mesure de la recherche traduit le degré de lisibilité des caractères isolés, son rapport avec la capacité de lecture va dépendre du degré de ressemblance entre l'épreuve de recherche et le mécanisme de la lecture et de l'importance des lettres comme aides à la lecture. Les lecteurs avertis disposent de diverses autres caractéristiques aux niveaux des mots, de la syntaxe et de la sémantique, pouvant les aider à comprendre le sens d'un texte imprimé. Pour s'en rendre compte, on n'a qu'à altérer la forme des mots en y faisant alterner des caractères de haut et de bas de casse ou en éliminant les espaces entre les mots par rapprochement ou par remplissage des espaces qui

les séparent. De telles modifications nuisent à la vitesse de lecture et de recherche lorsque l'objectif de la recherche vise un mot en particulier (Fisher, 1975; Spragins, Lefton et Fisher, 1976; et Fisher et Lefton, 1976). De plus, Malt et Seaman (1978) ont démontré que les améliorations de la capacité de lecture obtenues lorsqu'on utilisait une forme donnée de remplissage pour éliminer les espaces disparaissaient si l'on modifiait la forme du remplissage. La capacité de lecture tombait alors brusquement à son niveau initial. Il semble donc que le lecteur d'un texte dont les mots sont soudés le déchiffre lettre par lettre, sans qu'il soit prouvé pour autant que le même mécanisme joue dans la lecture d'un texte composé normalement. En fait, Wheeler (1970) suggère que les caractéristiques d'un niveau supérieur, comme celles que l'on trouve dans les digrammes et les trigrammes, aident probablement à reconnaître les mots. De la même façon, Dunn-Rankin (1978)signale que les combinaisons de lettres constituent des indications visuelles importantes pour la reconnaissance des mots, que les lettres soient réunies ou séparées. Cette conclusion découle d'analyses sur des groupes de mots classés par ressemblance. Cependant, on a découvert que la caractéristique visuelle prépondérante, lorsque l'on compare des mots, c'est la première lettre du mot.

Mason (1975) a étudié l'effet de la redondance spatiale sur l'identification des symboles littéraux. L'auteur a défini la redondance spatiale comme une corrélation entre les caractéristiques propres à une lettre et l'emplacement le plus fréquent de cette lettre dans les mots des textes anglais. Dans son expérience, les bons lecteurs étaient plus sensibles que les lecteurs faibles à la redondance spatiale de lettres séparées figurant dans des suites de lettres qui ne formaient pas de mots. De plus, chez les bons lecteurs, la recherche

des lettres cibles dans les anagrammes était plus rapide que dans les mots eux-mêmes quand les anagrammes avaient une plus grande redondance spatiale. Dans une autre expérience utilisant des symboles non alphabétiques au lieu de lettres (Mason et Katz, 1976) on a éliminé l'effet de confusion produit par la redondance séquentielle. Au cours de cette expérience, on a altéré la redondance spatiale en contrôlant la probabilité d'apparition d'un symbole à un endroit précis d'une suite de six symboles. Là encore, la redondance spatiale favorisait davantage les bons lecteurs que les lecteurs plus faibles, bien que les deux groupes eussent un rendement équivalent dans des conditions normales. Il semble donc que, pour les bons lecteurs, le fait qu'une lettre occupe une position attendue est un élément d'identification qui s'ajoute à la perception des caractéristiques visuelles. L'identification des lettres fait partie des automatismes de la lecture, en ce sens que l'emplacement d'une lettre n'a d'utilité qu'une fois cette lettre identifiée.

L'exposé qui précède ne cite pas toute la documentation utile à notre propos. Cependant, il suffit pour démontrer que la capacité d'extraire de l'information des lettres et des combinaisons de lettres dans un contexte familier fait partie des automatismes de la lecture. Par conséquent, si les effets de l'espacement des caractères et des lignes sur la facilité de différenciation des lettres sont importants, ils devraient aussi influer sur le processus de la lecture. Pour mettre cette hypothèse à l'épreuve, on pourrait examiner (a) les raisons expliquant pourquoi l'espacement peut avoir des effets différentiels sur la perception et (b) le mécanisme de la lecture.

Il existe certaines preuves psychophysiques à l'effet que la sensibilité fovéale n'est pas indépendante de ce qui se produit dans le champ visuel. Par

exemple, Breitmeyer et Valberg (1979) ont découvert que la région fovéale est moins sensible à une stimulation lumineuse lorsque le champ visuel perçoit un réseau de signaux oscillants. De plus, Mackworth (1965) signale que si l'on ajoute à un affichage des lettres qui n'ont rien à voir avec les mots, le sujet de l'expérience avait beaucoup de difficultés à comparer avec justesse une lettre présentée à la fovéa avec deux lettres présentées dans le champ visuel. Mackworth (1965) a recours au concept du "champ visuel utile" pour interpréter ses constatations. Le champ visuel utile est centré sur le point de fixation de l'oeil. Lorsque l'oeil reçoit un trop grand nombre d'informations, le champ visuel utile se contracte pour éviter une surcharge du système visuel. Ainsi, lorsqu'on réduit l'espacement entre les caractères et les lignes, la superficie du champ visuel utile devrait se contracter et ralentir, de ce fait, la vitesse de balayage au cours de la lecture. Le ralentissement de la vitesse de balayage résulterait aussi bien de la réduction de l'information fournie par le champ visuel que de la réduction de la sensibilité fovéale. Cependant, l'augmentation du nombre de caractères occupant une superficie donnée compenserait probablement l'effet de la réduction de l'information fournie par le champ visuel.

Une épreuve visant à étudier l'espacement des caractères d'une page de texte devrait ressembler à une épreuve de lecture portant sur plusieurs lignes de caractères. D'après le mécanisme normal de la lecture d'un texte écrit en anglais, l'épreuve devrait amener le point de fixation de l'oeil à se déplacer de gauche à droite sur la ligne de caractères. Si on se servait du même texte pour une épreuve de recherche, on pourrait constater des variations dans la vitesse de balayage le long d'une seule ligne, en termes de distance parcourue comme en termes de caractères lus. Il se peut que les caractères

soient lus plus lentement lorsqu'ils sont plus distancés parce que l'oeil doit balayer une plus grande distance. Cependant, la vitesse de balayage, en termes de temps de parcours de la distance couverte pour lire chaque unité, devrait s'améliorer si l'on augmente l'espacement des caractères, afin de conserver la vitesse des caractères. Le temps de parcours qu'exige chaque unité devrait s'améliorer jusqu'à ce que l'espacement horizontal des caractères soit suffisamment compensé par la vitesse de déplacement de l'oeil pour donner au lecteur le maximum d'information juste avec le minimum d'effort. Par définition, il devrait y avoir corrélation entre ce phénomène et le plus bas taux d'erreurs.

Description de l'expérience

On a présenté à chacun des dix sujets participant à l'expérience des images comme celle qui est reproduite à la figure 1.

Intercaler la figure 1 à peu près à cet endroit

La photographie de la figure 1 représente un écran de télévision et montre le jeu de caractères 5 sur 7 utilisés. Un algorithme d'ordinateur a choisi au hasard les lettres affichées parmi toutes les lettres minuscules de l'alphabet. Un signe "plus" (+) dans la marge de gauche indiquait la ligne sur laquelle on demandait au sujet de porter son attention. Nous avons choisi comme cibles de notre recherche les cinq lettres descendantes (g, j, p, q et y), parce que leur longueur constituait une variable intéressante. Une de ces cinq lettres se trouvait toujours au début de la ligne à balayer et indiquait au sujet la lettre cible au début de chaque essai. Le sujet devait alors balayer la ligne pour y repérer cette lettre et, dès qu'il la rencontrait, nommer immédiatement la lettre suivante. Dans la figure où l'on visait la lettre "q", il devait nommer la lettre "v". On demandait en outre au sujet de balayer la ligne une seule fois. S'il arrivait à la fin de la ligne sans avoir pu identifier la lettre cible, on lui demandait de nommer la dernière lettre de la ligne. La lettre cible, de même que la lettre qui la suivait et la dernière lettre de la ligne, n'apparaîssaient qu'une seule fois dans la ligne après l'indication donnée par la première lettre. On demandait aussi au sujet d'essayer de

régulariser sa vitesse de balayage de façon à ne pas rater plus de 10 p. cent de ses essais. L'image apparaîssait de façon relativement instantanée et la réponse donnée actionnait un interrupteur commandé par la voix qui effaçait l'image. On mesurait le temps de réponse à la milliseconde (ms) près et l'expérimentateur introduisait la réponse proprement dite au pupitre de commande pendant les trois secondes d'intervalle entre chaque essai.

La matrice des caractères 5 sur 7 embrassait horizontalement 10 éléments imagiers et verticalement, 14 lignes d'exploration télévisuelle afin de réduire au minimum le scintillement qui résulterait de l'affichage de chaque point de la matrice sur seulement une trame du téléviseur. Au cours de l'expérience, nous avons utilisé un écran-témoin de télécouleur de 19 pouces de marque Conrac, modèle RHN 19/C, muni d'éléments circulaires de substance luminescente disposés selon le modèle triangulaire habituel. Les lettres étaient toujours affichées en blanc sur fond neutre. Nous avons calibré la couleur blanche à la température de couleur de 6500^{o}K, et réglé à 14 lamberts-pied l'intensité de l'image lorsqu'elle était complètement éclairée; l'éclairage ambiant de la pièce se situait à environ 34 candélas-pied.

Le sujet s'assoyait à environ quatre pieds de l'écran, la tête non retenue. L'expérimentateur affichait une page pour mieux expliquer l'épreuve au sujet, puis il lui faisait faire 10 essais préliminaires. L'expérience en soi consistait en un plan de 27 mesures répétées, tenant compte de trois facteurs: la longueur des descendantes, l'espacement horizontal et l'espacement vertical. Par longueur des descendantes, on entend le degré de dépassement des caractères descendants sous la ligne, soit 0, 1 ou 2 éléments imagiers. L'espacement vertical et l'espacement horizontal étaient de 1, 2 ou 3 éléments imagiers.

Le plan total comprenait donc 27 cellules. Chacune des cinq lettres cibles était présentée à deux reprises dans chaque condition d'expérimentation. Les 27 conditions apparaissaient au hasard dans chacun des 10 groupes d'essais. Le nombre total d'essais par sujet se chiffrait donc à 270. Étant donné que la session d'expérimentation durait environ 35 minutes, le sujet avait droit à une courte pause après le 135^{e} essai.

Nous avons effectué une analyse de variance des mesures répétées sur chacune des trois variables dépendantes suivantes: (1) la fréquence des erreurs, (2) la vitesse de balayage exprimée en caractères à la seconde, et (3) la vitesse de balayage exprimée en éléments imagiers à la seconde. L'analyse de la fréquence des erreurs a révélé un effet marqué de l'espacement horizontal ($F(2,18)=5.598$, $p<.05$) et un certain effet de la longueur des descendantes ($F(2,18)=15.944$, $p<.01$). Aucune interaction n'était significative. L'analyse de la vitesse de balayage exprimée en caractères à la seconde a révélé un effet marqué de l'espacement vertical ($F(2,18)=5.374$, $p<.05$), de l'espacement horizontal ($F(2,18)=42.720$, $p<.01$) et de la longueur des descendantes ($F(2,18)=15.745$, $p<.01$). Aucune interaction n'était significative non plus. L'analyse de la vitesse de balayage exprimée en éléments imagiers à la seconde a révélé un effet marqué de l'espacement vertical ($F(2,18)=5.60$, $p<.05$), de l'espacement horizontal ($F(2,18)=205.325$, $p<.01$) et de la longueur des descendantes ($F(2,18)=18.321$, $p<.01$). Les figures 2, 3 et 4 indiquent les moyennes correspondant à ces effets. Le tableau 1 répartit la valeur de la variance totale pour chacun des effets marqués.

Intercaler les figures 2, 3 et 4 à peu près à cet endroit

Intercaler le tableau 1 à peu près à cet endroit

Nous avons vérifié la signification statistique des écarts entre les moyennes, à l'aide de la procédure Newman-Keuls (Winer, 1962). Le fait de ne pas faire dépasser les caractères descendants sous la ligne a entraîné beaucoup plus d'erreurs et a donné une vitesse de balayage beaucoup plus lente que lorsque l'on faisait dépasser les descendantes d'un élément imagier ou deux. On ne notait cependant pas de différence importante selon que le dépassement était d'un élément imagier ou de deux. On peut déduire de ces constatations que la queue d'un caractère descendant devrait dépasser la ligne d'au moins un élément imagier.

La raison de l'effet de la longueur des descendantes est probablement fort simple: une descendante qui dépasse la ligne, même d'un seul élément imagier, est assez remarquable pour en faciliter la justesse et la rapidité de détection, compte tenu du fait que la probabilité d'apparition d'un caractère descendant est a priori inférieure à 0.2.

L'importance de l'espacement entre les lignes de caractères n'a pas semblé modifier la fréquence des objectifs ratés, mais a eu un effet sur la vitesse de balayage de la ligne. En fait, un espacement vertical de trois éléments imagiers a produit un balayage de la ligne sensiblement plus rapide qu'un espacement d'un élément imagier ou deux. La différence était insignifiante selon que l'espacement était d'un élément imagier ou de deux. Ces constatations nous donnent une certaine latitude pour établir quel devrait être l'espacement minimum. Ainsi, si la vitesse de balayage importe peu, l'espacement minimum

d'un élément imagier semble suffisant. Cependant, si l'on a comme critère un balayage de ligne plus rapide, il faudrait conserver un espacement d'au moins trois éléments imagiers entre les lignes. Si l'on a obtenu un balayage plus rapide parce que l'on avait moins d'efforts à faire pour identifier les caractéristiques des lettres, c'est peut-être que si les lignes d'un texte sont séparées par au moins trois éléments imagiers, elles sont moins fatigantes à lire que si elles sont séparées par un élément imagier ou deux.

L'effet de l'espacement entre les lignes sur la vitesse de balayage horizontal ne s'explique pas de façon évidente à première vue. Il est possible que la diminution des interférences visuelles dans le champ de vision, qui résultent de l'augmentation de l'espacement entre les lignes, agrandisse le "champ visuel utile" (Mackworth, 1965). Si on agrandit le champ visuel utile au moins dans le sens horizontal, le supplément d'information puisé dans cette partie du champ visuel qui se trouve dans le sens du balayage pourrait donner des vitesses de balayage plus rapides.

En espaçant d'un élément imagier les caractères d'une ligne (espacement horizontal) on a obtenu beaucoup plus de fautes qu'en les espaçant de deux ou trois éléments. La différence était insignifiante entre ces deux derniers espacements quant au nombre de lettres cibles non repérées. Les analyses effectuées sur les deux mesures de la vitesse de balayage ont révélé des différences marquées entre ces trois espacements. Cependant, nous avons constaté qu'en augmentant l'espacement, le nombre de caractères balayés à la seconde diminuait, alors que le nombre d'éléments imagiers balayés à la seconde augmentait.

Il y a lieu de croire que nous avons obtenu le taux d'assimilation de caractères le plus élevé avec l'espacement le plus réduit, parce que nous avions concentré les caractères en question dans une unité de distance. C'est probablement à cette forte densité de caractères qu'est aussi attribuable la fréquence élevée de fautes obtenue à la même occasion; en effet, la qualité a baissé légèrement, mais de façon notable, à la vitesse la plus rapide de balayage des caractères.

L'augmentation de l'espacement entre les caractères permettrait de percevoir un champ visuel plus étendu au cours d'une période de temps donné. Cette perception plus vaste peut, elle aussi, être imputable à l'agrandissement du champ visuel utile résultant de la moins grande densité des lettres. Par ailleurs, la perception d'un champ visuel plus étendu dans le temps de parcours de chaque unité balayée pourrait être imputable à la plus grande rapidité de déplacement de l'oeil et au raccourcissement du temps de fixation. Cependant, l'augmentation de la vitesse de perception spatiale n'a pas suffi à conserver la vitesse de lecture des caractères observée avec l'espacement le plus réduit.

Pour établir quel serait le meilleur espacement horizontal, on doit tenir compte des trois mesures dépendantes. La mesure de la fréquence des erreurs et celle de la vitesse donnent à croire que l'on devrait espacer les caractères d'au moins deux éléments imagiers pour que les lettres soient identifiées le plus rapidement et le plus justement possible. Un espacement plus grand encore augmenterait sensiblement la vitesse de perception spatiale, mais réduirait néanmoins le taux d'assimilation des caractères. C'est là un point important parce que l'information provient des caractères, et non, en soi, de la distance parcourue lors de la lecture.

L'effet de l'espacement horizontal sur la vitesse de balayage des caractères peut être attribuable au changement de densité des caractères dont s'accompagne la modification de l'espacement horizontal. Bien que la forte densité de l'espacement d'un seul élément imagier ait semblé exiger une vitesse de lecture trop élevée, comme l'indique l'augmentation de la fréquence des erreurs qui en a résulté, la densité de l'espacement de deux éléments imagiers n'a pas semblé trop forte. Si cela avait été, l'augmentation importante de la vitesse de perception de l'espace ne se serait pas produite. De plus, la fréquence des erreurs commence à devenir asymptotique avec l'utilisation de l'espacement horizontal de deux éléments imagiers. Il est donc probable que l'espacement horizontal minimum de deux éléments imagiers, que nous recommandons, soit effectivement dû aux limites de la vitesse de transmission de l'information chez l'homme. Cette conclusion doit inciter à la prudence lorsqu'il s'agit d'appliquer l'espacement horizontal recommandé par notre expérience à une page de texte qui présentera forcément des redondances et des caractéristiques d'un ordre supérieur. Au cours de la lecture, le sujet analyse l'information perçue au niveau de la lettre, au niveau du mot et au niveau des locutions. L'effet de l'espacement d'un seul élément imagier entre les caractères peut être impossible à déceler sur les erreurs de lecture dans le contexte de toutes ces données additionnelles.

Conclusions de l'expérience

A la lumière des données obtenues au cours de notre expérience sur la vitesse et sur la justesse, nous pouvons recommander l'espacement minimum requis entre les caractères et entre les lignes, de même que la longueur minimale

requise pour le dépassement de la queue des caractères descendants. Nos recommandations s'appliquent à la présentation d'un texte sur un écran de télévision dont le lecteur se trouve éloigné d'environ quatre pieds, dont les caractères sont représentés sur une matrice de 5 points sur 7 embrassant verticalement 14 lignes d'exploration télévisuelle, et dont les éléments de substance luminescente sont ronds plutôt qu'allongés. Nos recommandations présument que les écarts obtenus dans les fréquences d'erreurs ont une signification pratique aussi bien que statistique, et que l'identification des lettres constitue un élément important de l'automatisme de la lecture. Nos recommandations sont les suivantes:

a) Espacement vertical:

Les résultats de notre expérience montrent que l'espacement entre les lignes n'influe pas sur la justesse, même quand il n'est que d'un seul élément imagier. Cependant, on peut améliorer sensiblement la vitesse d'identification des lettres, et possiblement rendre la lecture plus facile, en augmentant l'espacement entre les lignes à trois éléments imagiers. Il est donc recommandé d'espacer les lignes soit d'un, soit de trois éléments imagiers selon que la vitesse de lecture constitue un critère important ou non. Cette recommandation ne tient pas compte de la longueur choisie pour la queue des caractères descendants. L'espacement vertical de trois éléments imagiers offre en outre l'avantage de fournir l'espace nécessaire pour les accents dans une langue comme le français.

b) Espacement horizontal:

Nous recommandons un espacement de deux éléments imagiers entre les caractères d'une même ligne. Un espacement moindre augmente sensiblement le taux d'erreurs dans l'identification des lettres, alors qu'un espacement supérieur réduit inutilement la vitesse d'entrée de l'information.

c) Longueur des descendantes:

Nous recommandons que la queue d'un caractère descendant dépasse la ligne d'au moins un élément imagier. Le lecteur identifie plus lentement et avec beaucoup plus d'erreurs les caractères descendants dont la queue ne descend pas sous la ligne. Un dépassement supérieur à un élément imagier n'améliore pas sensiblement le rendement.

Ouvrages de référence

Breitmeyer, B. G. et Valberg, A. Local foveal inhibitory effects of global peripheral excitation. Science, 1979, 203: 463-464.

Breland, K. et Breland, M. K. Legibility of newspaper headlines printed in capitals and in lower case. Journal of Applied Psychology, 1944, 28: 117-120.

Dunn-Rankin, P. The visual characteristics of words. Scientific American, 1978, 238: 122-130.

Fisher, D. F. Reading and visual search. Memory and Cognition, 1975, 3: 188-196.

Fisher, D. F. et Lefton, L. A. Peripheral information extraction: A developmental examination of reading processes. Journal of Experimental Child Psychology, 1976, 21: 77-93.

Mackworth, N. H. Visual noise causes tunnel vision. Psychonomic Science, 1965, 3: 67-68.

Malt, B. C. et Seaman, J. G. Peripheral and cognitive components of eye guidance in filled-space reading. Perception and Psychophysics, 1978, 23: 399-402.

Mason, M. Reading ability and letter search time: Effects of orthographic structure defined by single-letter positional frequency. Journal of Experimental Psychology: General, 1975, 104: 146-166.

Mason, M. et Katz, L. Visual processing of nonlinguistic strings: Redundancy effects and reading ability. Journal of Experimental Psychology: General, 1976, 105: 338-348.

Paterson, D. G. et Tinker, M. A. Readability of newspaper headlines printed in capitals and lower case. Journal of Applied Psychology, 1946, 30: 161-168.

Snyder, H. L. et Maddox, M. E. Information transfer from computer-generated dot-matrix displays. Report HFL-78-3/ARO-78-1, U.S. Army Research Office, Research Triangle Park, N.C., 1978.

Spragins, A. B., Lefton, L. A. et Fisher, D. F. Eye movements while reading and searching spatially transformed text: A developmental examination. Memory and Cognition, 1976, 4: 36-42.

Tinker, M. A. Examiner's manual for Tinker Speed of Reading Test. Minneapolis: University of Minnesota Press, 1955.

Tinker, M. A. Bases for Effective Reading. Minneapolis: University of Minnesota Press, 1965.

Warren, A. L. The Perceptibility of Lower Case and All Capitals Newspaper Headlines. Thèse de maîtrise (1942) résumée dans Cornog, D. Y. et Rose, F. C. Legibility of Alphanumeric Characters and Other Symbols: II. A reference Handbook. Washington: U.S. Government Printing Office, National Bureau of Standards Miscellaneous Publication 262-2.

Wheeler, D. D. Process in word recognition. Cognitive Psychology, 1970, 1: 59-85.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Tableau 1

Pourcentage de variance totale de chaque mesure, pour chacune des variables indépendantes

Variable indépendante	Variables dépendantes		
	Fréquence des erreurs	Caractères à la seconde	Eléments imagiers à la seconde
Espacement vertical	---	1,3	1,1
Espacement horizontal	4,3	6,6	16,6
Longueur des descendantes	7,2	3,2	3,2

Figure 1.

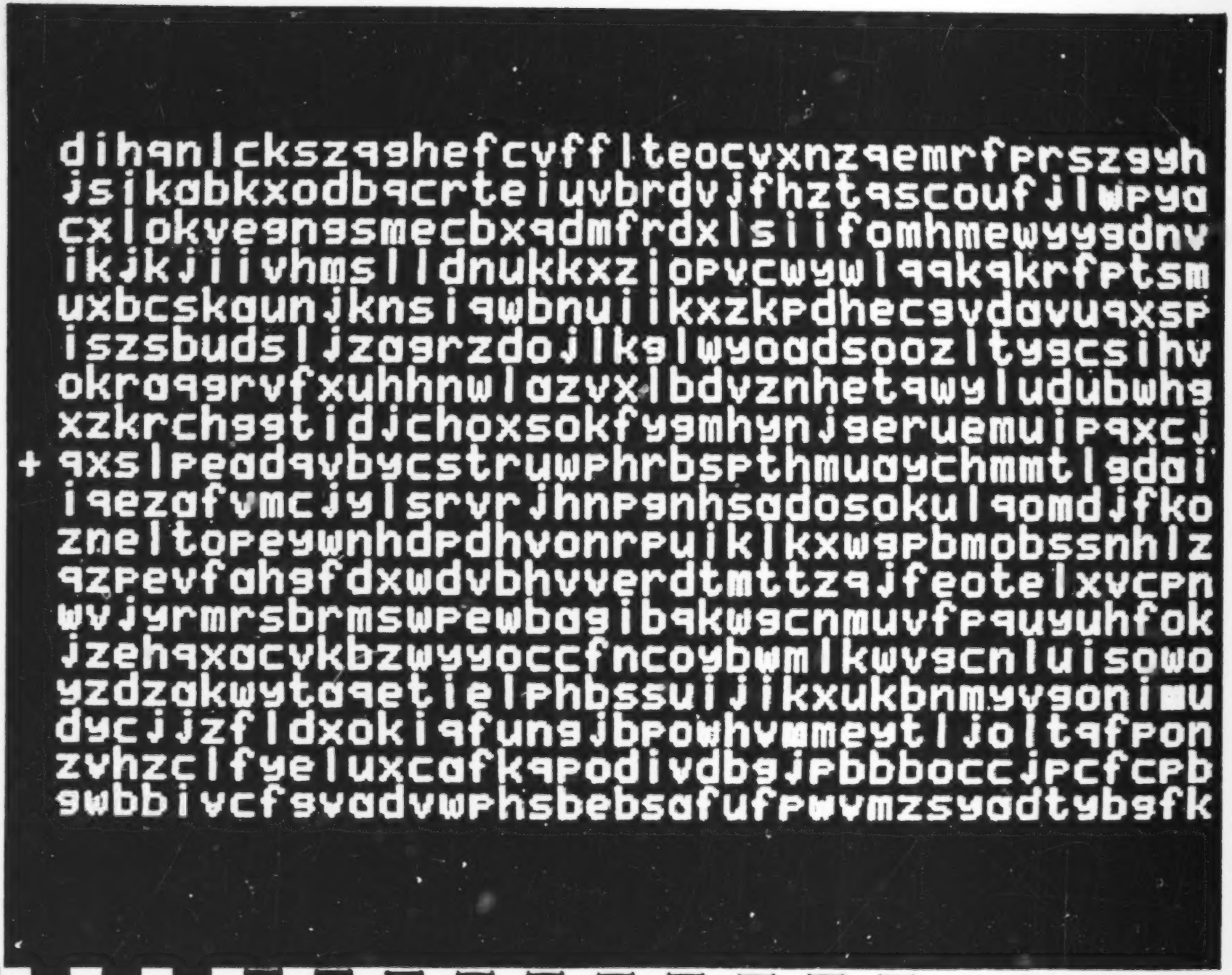

dihqnlckszqghefcvfflteocvxnzqemrfprszgyh
jsikobkxodbqcrteiuvbrdvjfhztqscoufjlwpyo
cxlokvegngsmecbxqdmfrdxlsiifomhmewyygdnv
ikjkjjivhmsslldnukkxziopvcwywlqqkqkrfptsm
uxbcskounjknsiqwbnuiikxzkpdhecgvdovuqxsp
iszsbudsljzogrzdojlkglwyoodsoozltygcsihv
okroqgrvfxuhhnwlozvxlbdvznhetqwyludubwhg
xzkrchggtidjchoxsokfygmhynjgeruemuipqxcj
+ qxslpeodqvbycstruwphrbspthmuoychmmtlgdoi
iqezofvmcjylsrvrjhnpgnhsodosokulqomdjfko
znetopeywnhdpdhvonrpuiklkxwgpbmobssnhlz
qzpevfohgfdxwdvbhvverdtmttzqjfeotelxvcpn
wvjyrmrsbrmswpewbogibqkwgcnmuvfpquyuhfok
jzehqxocvkbzwyyoccfncoybwmlkwvgcnluisowo
yzdzokwytoqetielphbssuijikxukbnmyygonimu
dycjjzfldxokiqfungjbpowhvmmeytljoltqfpon
zvhzclfyeluxcofkqpodivdbgjpbbboccjpcfcpb
gwbbivcfgvodvwphsbebsofufpwvmzsyodtybgfk

Figure 2:

Rapport entre l'espacement des lignes de caractères et la vitesse de balayage lors de la recherche de la lettre cible.

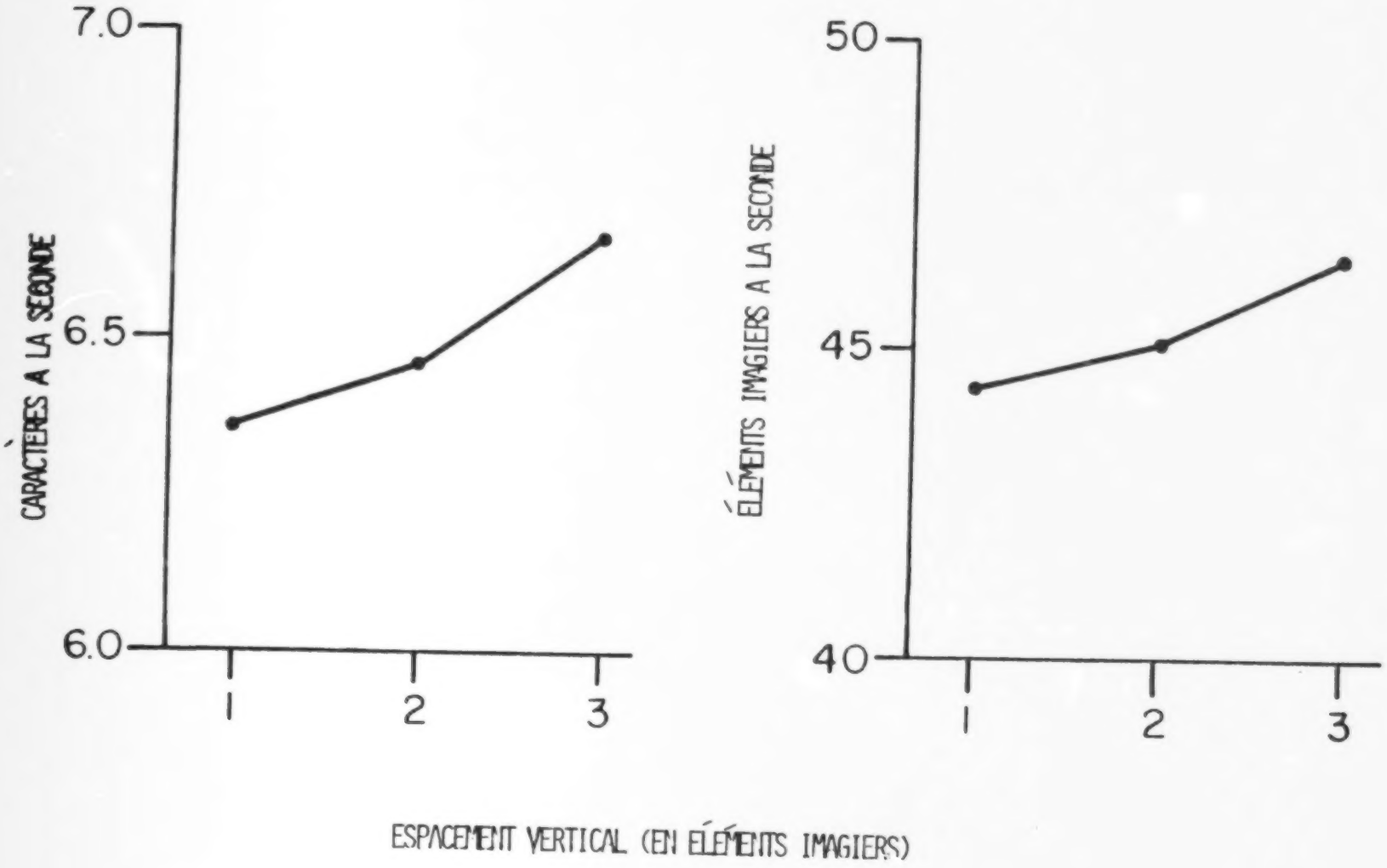


Figure 3:

Fréquence des erreurs et vitesse de balayage lors de la recherche de la lettre cible, exprimées en fonction de l'espacement des caractères d'une ligne.

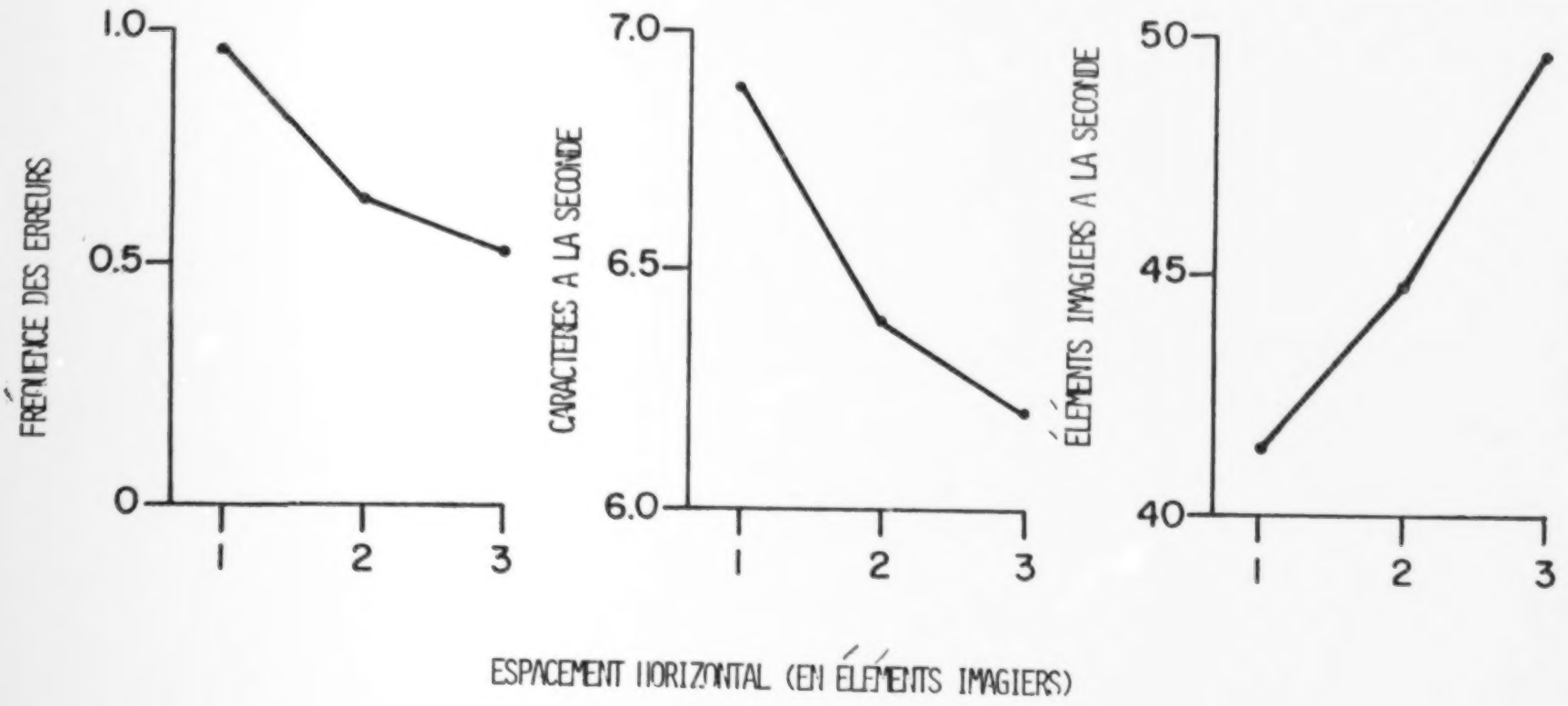
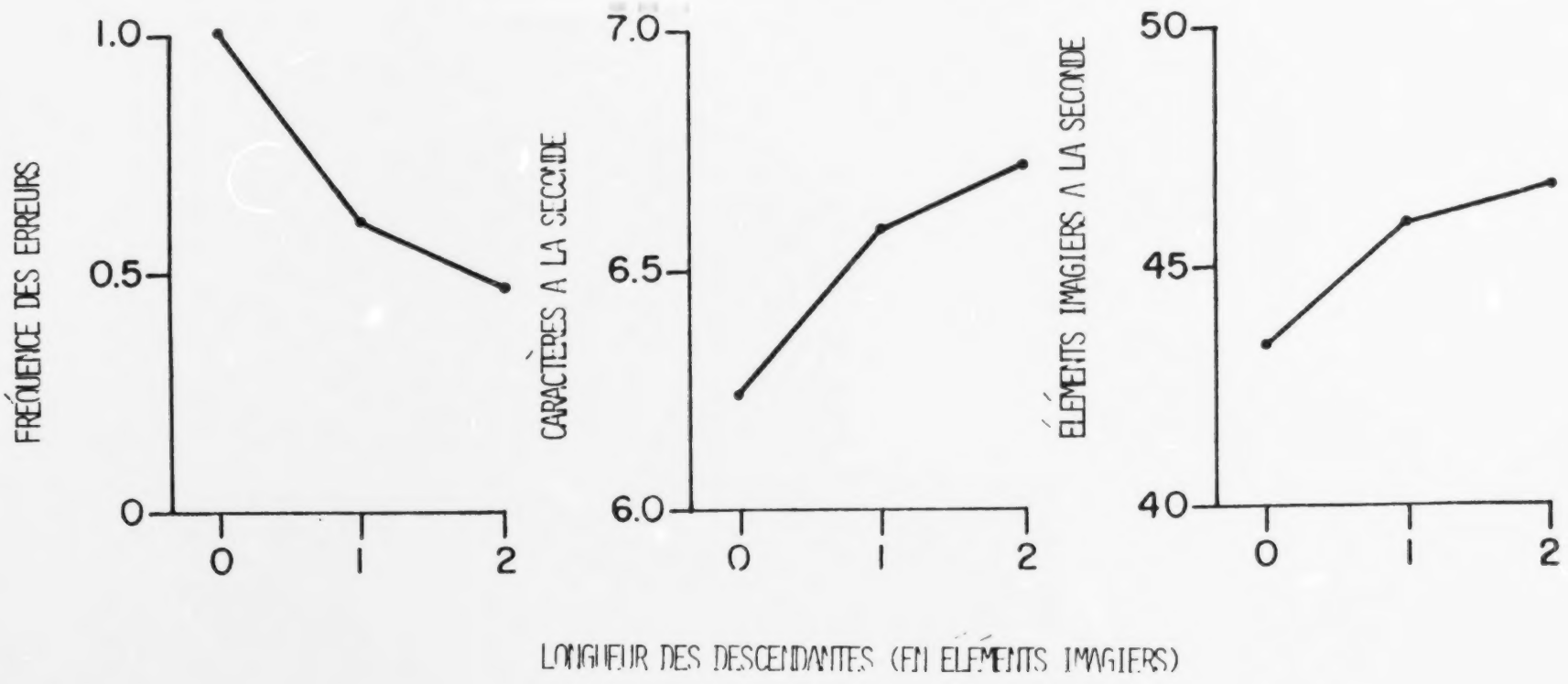


Figure 4:

Fréquence des erreurs et vitesse de balayage lors de la recherche de la lettre cible, exprimées en fonction du nombre d'éléments imagiers utilisés pour le dépassement des descendantes.



Jeux de caractères des terminaux vidéotex servant aux expériences-pilotes

W.C. Treurniet et C.D. O'Brien

Centre de recherches sur les communications

Ottawa (Ontario)

Novembre 1979

Résumé

Le Groupe de recherche en comportement du ministère des Communications a étudié un certain nombre de paramètres des jeux de caractères, dans le but de choisir les jeux de caractères les mieux appropriés pour l'entrée en Rouge, Vert et Bleu (RVB) et pour l'entrée vidéo-composée. Le Groupe a choisi deux jeux de caractères pour l'anglais et deux pour le français, qui serviront aux spécifications des terminaux utilisés pour les expériences-pilotes. Pour les espacements, le Groupe a proposé une solution quasi optimale qui tient compte à la fois des facteurs de lisibilité et des impératifs économiques motivant le choix des concepteurs. Le Groupe a également recommandé que l'on prescrive l'espacement proportionnel et l'augmentation de l'espacement des caractères lorsqu'on adopte le mode de composition libre de textes avec adressage direct aux mots.

1. Jeux de caractères pour les terminaux servant aux expériences-pilotes

Le Groupe de recherche en comportement du ministère des Communications a étudié (Treurniet, 1979a; Treurniet, 1979b; et Treurniet et Latrémouille, 1979) la forme et l'organisation des caractères d'une page. Le Groupe avait pour but de définir les paramètres du système qui donneraient la meilleure lisibilité possible d'un texte sur l'écran de télévision ordinaire que l'on utiliserait au cours des expériences-pilotes menées sur le système vidéotex canadien. On a constaté que, selon qu'on adoptait une technique ou une autre, on avait besoin de jeux de caractères et d'espacements différents pour pallier différentes limites du système.

On dit qu'un texte est transparent quand on peut oublier son apparence et sa forme aussitôt après en avoir extrait la signification. Inversement, tout ce qui gêne nos habitudes de lecture donne une certaine opacité au texte (Roudiez, 1978). A moins que cette opacité ne soit voulue (comme dans certaines formes d'expression artistique), il vaut mieux la réduire au minimum afin que le lecteur puisse concentrer toute son attention sur la compréhension du texte. Entre autres choses, l'opacité peut provenir de lettres, de mots ou de phrases d'une page dont la présentation est inattendue ou inesthétique. Les pages de texte du système vidéotex devraient donc être imprimées avec un jeu de caractères qui pallie dans la mesure du possible les limites inhérentes à ce support.

Plusieurs expériences et démonstrations ont permis d'établir les jeux de caractères les plus adéquats pour les deux modes d'entrée vidéo proposés pour les systèmes vidéotex. Les signaux vidéo-composée. et les signaux vidéo RVB sont deux sortes de signaux vidéo dont les effets sur l'information affichée sont nettement différents. Par rapport au signal d'entrée RVB, le signal vidéo-composée dégrade l'image en limitant la résolution et la reproduction des coleurs. Le changement de couleur d'une zone à l'autre produit du scintillement sur la ligne de rencontre verticale de ces deux zones. Il est difficile d'empêcher l'effet du scintillement à cause de la façon de décoder la couleur du signal vidéo-composée. Les limites imposées par les normes de télévision NTSC sur la largeur de bande empêchent les courts changements de luminance requis pour afficher adéquatement des lignes verticales larges d'un élément imagier. Pour les images RVB chromatiques et achromatiques et pour les images vidéo-composée achromatiques, nous recommandons les jeux de caractères CSA-1 (anglais) et CSA-2 (français) équivalents à la norme ISO646; ces jeux sont décrits ci-dessous et illustrés dans les figures des annexes. Avec le jeu de caractères CHSET 7 décrit ci-dessous, on doit adopter l'espacement proportionnel, que nous recommandons fortement pour tous les jeux de caractères utilisés dans le mode de composition libre de textes avec adressage direct aux mots.

On peut faire des compromis économiques et techniques entre la façon de construire un terminal et les exigences particulières du meilleur jeu de caractères. Dans les descriptions que nous donnons ci-dessous, nous présentons les exigences primordiales du point de vue de la lisibilité. Dans une autre section, nous tenons compte de ces compromis et recommandons des exigences réalisables en pratique pour les terminaux servant aux expériences-pilotes.

ENTREE VIDEO RVB

CHSET 1: Ce jeu de caractères CSA-1 utilise une matrice de caractères 5 sur 8. Il permet d'afficher sur chaque page 20 lignes de 40 caractères. La ligne inférieure d'éléments imagiers de la matrice ne sert qu'aux caractères descendants. Nous recommandons d'afficher les textes avec un espacement horizontal de deux éléments imagiers et un espacement vertical de trois éléments imagiers entre les matrices de caractères.

CHSET 2: Ce jeu de caractères CSA-2 utilise une matrice de caractères 5 sur 11 permettant d'afficher sur chaque page 20 lignes de 40 caractères. La ligne inférieure d'éléments imagiers de la matrice ne sert qu'aux caractères descendants, alors que les trois lignes supérieures assurent l'espacement vertical avec la ligne du dessus. Ces trois lignes supérieures d'éléments imagiers de la matrice servent à accentuer 10 caractères majuscules alors que les deuxième, troisième et quatrième lignes servent à accentuer 10 minuscules. Nous recommandons d'afficher le texte avec un espacement horizontal de deux éléments imagiers entre les matrices de caractères et de ne laisser aucun espacement vertical entre les matrices.

CHSET 3: Ce jeu de caractères CSA-1 utilise une matrice de caractères 7 sur 11. Il permet d'afficher sur une page 16 lignes de 32 caractères. Les deux lignes inférieures d'éléments imagiers de la matrice ne servent qu'aux caractères descendants. Nous recommandons d'afficher le texte avec un espacement horizontal de deux éléments imagiers et un espacement vertical d'un élément imagier entre les matrices de caractères.

CHSET 4: Ce jeu de caractères CSA-2 est identique au jeu CHSET 3, sauf que des minuscules accentuées remplacent 10 caractères non alphabétiques. On utilise les deux lignes supérieures de la matrice pour les accents. Nous recommandons d'afficher le texte avec un espacement horizontal de deux éléments imagiers et un espacement vertical d'un élément imagier entre les matrices.

ENTREE VIDEO-COMPOSITE

CHSET 5: Ce jeu de caractères CSA-1 ressemble au jeu CHSET 3, sauf que la largeur de toutes les composantes des caractères, de même que les espaces à l'intérieur des caractères, ont au moins deux éléments imagiers de largeur. La largeur excédentaire pallie la restriction de la largeur de bande du signal vidéo-composée. La plupart des caractères de ce jeu n'utilisent que six des sept colonnes de la matrice de caractère. En dépit de la restriction de la largeur de bande, nous recommandons donc d'afficher le texte avec un

espacement horizontal de deux éléments imagiers et un espacement vertical d'un élément imagier entre les matrices.

CHSET 6: Ce jeu de caractères CSA-2 ressemble au jeu CHSET 5 en ce sens que la largeur de toutes les composantes des caractères compte au moins deux éléments imagiers. Les caractères accentués ressemblent à ceux du jeu CHSET 4. Nous recommandons d'afficher le texte avec un espacement horizontal de deux éléments imagiers et un espacement vertical d'un élément imagier entre les matrices.

CHSET 7: Ce jeu de caractères a été créé pour améliorer l'apparence de deux lettres des jeux CHSET 5 et CHSET 6, dont les sept colonnes ne permettaient pas d'étaler suffisamment les composantes verticales des lettres m et w. On a donc conçu le jeu CHSET 7 avec matrice 10 sur 11. Pour les lettres m et w, majuscules et minuscules, on utilise les dix colonnes de la matrice. Tous les autres caractères n'utilisent que six colonnes, comme dans les jeux CHSET 5 et CHSET 6. Etant donné que la majorité des caractères du jeu CHSET 7 laissent inutilisées quatre colonnes de la matrice, on doit supprimer ces vides au moyen de l'espacement proprotionnel lors de l'affichage du texte.

CHSET 8: Ce jeu de caractères CSA-2 utilise une matrice de caractère 5 sur 9 permettant d'afficher 20 lignes de 40 caractères sur une page. La ligne inférieure des éléments imagiers de la matrice ne sert qu'aux caractères descendants, alors que les deux lignes supérieures servent à accentuer 10 caractères minuscules. Nous recommandons d'afficher le texte avec un espacement horizontal et vertical de deux éléments imagiers entre les matrices.

Ces jeux de caractères impliquent trois niveaux d'exploitation du système. Au premier niveau, le terminal ne donne que les images complètes (vidéo-composite); dans ce cas, on devrait utiliser les jeux CHSET 5 et CHSET 6. Au deuxième niveau, les signaux de sortie RVB sont disponibles pour les téléviseurs équipés pour les recevoir; avec cette structure, on pourrait utiliser les jeux CHSET 1, CHSET 2, CHSET 3, CHSET 4, et CHSET 8. Le troisième niveau permettrait l'espacement proportionnel. Lorsque l'on utilise des terminaux pouvant produire des images complètes, on devrait utiliser le jeu CHSET 7 en mode de composition libre de textes.

Pour les affichages RVB, on peut utiliser les couleurs qui ont un bon contraste de luminance. Cependant, nous ne recommandons pas l'utilisation de la couleur pour afficher des textes sur des écrans vidéo-composites, parce qu'il en résulte une grande instabilité de l'image.

2. Recommandations tenant compte des facteurs économiques et techniques

Il est généralement convenu, à l'échelle internationale, qu'une ligne de texte devrait compter 40 caractères, ce qui s'est révélé possible sur les systèmes RVB seulement. Les systèmes vidéo-composée requièrent des caractères plus gros à cause de la restriction de la largeur de bande.

A cause du nombre réduit de lignes de balayage disponibles sur un signal de télévision NTSC, il faut utiliser moins que les 24 ou 25 lignes de texte proposées pour les systèmes de télévision européens. L'idéal serait d'utiliser un maximum de 20 lignes. Avec l'entrée vidéo-composée, on devrait utiliser un jeu de caractères proportionnellement une fois et demie plus grands (16 lignes de 32 caractères).

Les études de comportement ont montré que l'espacement horizontal entre les caractères 5 sur 7 devrait être de deux éléments imagiers, que la hauteur de la matrice devrait s'étendre à huit éléments imagiers pour recevoir les descendantes, et que trois autres éléments imagiers devraient séparer les lignes de caractères adjacentes. On pourrait ainsi afficher 20 lignes de 40 caractères s'il y avait 61 600 (220 sur 280) points adressables sur l'écran. De la même façon, l'espacement horizontal entre les caractères 7 sur 9 devrait être d'au moins deux éléments imagiers, la hauteur de la matrice devrait s'étendre à 11 éléments imagiers pour recevoir les descendantes, et un élément imagier supplémentaire devrait séparer les lignes de caractères adjacentes. On pourrait ainsi afficher 16 lignes de 32 caractères sur l'écran, avec 55 296 (192 sur 288) points adressables. Par conséquent, afin d'utiliser les meilleurs espacements entre les caractères et de tenir compte du surbalayage de la télévision, on doit exploiter la capacité d'affichage complète de 76 800 (240 sur 320) points adressables pour afficher les textes en respectant les densités maximales faisant l'objet d'ententes internationales sur d'autres bases.

Toutefois, on a décidé, pour des raisons d'économie, de limiter à 200 éléments imagiers verticaux et 256 éléments horizontaux la mémoire de visualisation par profil de multiplets. Cette décision exige qu'on modifie le nombre maximum de caractères qu'on affichera sur une page, ou l'espacement entre les caractères. C'est à cause du coût de la mémoire de visualisation qu'on a réduit la zone d'affichage disponible. Étant donné que le coût des mémoires diminue continuellement, cette restriction technique peut être relativement temporaire. On a donc opté pour un espacement quasi optimal mais acceptable des caractères, préférant ce compromis à une réduction du nombre de caractères par ligne.

Pour répondre aux exigences de la densité des textes, nous avons choisi les jeux de caractères CHSET 1, CHSET 8, CHSET 5 et CHSET 6 en guise de spécifications pour les terminaux servant aux expériences-pilotes. Le jeux CHSET 1 comptera un espacement horizontal d'un élément imagier et un espacement vertical de deux éléments entre les matrices. Le jeu CHSET 8 aura le même espacement horizontal, mais un espacement vertical d'un élément imagier seulement vu que l'autre élément est inclus dans la matrice de caractère pour faciliter l'accentuation. Les jeux CHSET 1 et CHSET 8 sont destinés à servir uniquement aux systèmes RVB.

Les jeux CHSET 5 et CHSET 6 sont destinés aux systèmes RVB et aux systèmes vidéo-composée. A cause des restrictions imposées sur la zone d'affichage, l'espacement horizontal et vertical de ces jeux sera d'un élément imagier seulement.

L'obligation de limiter à 40 le nombre de caractères par ligne ne s'applique qu'aux textes de commentaires, dans lesquels les caractères doivent conserver un rapport spatial fixe avec le reste de l'image. La composition libre de textes avec adressage direct aux mots offre beaucoup plus de souplesse. Avec ce mode de composition, le terminal peut effectuer l'espacement proportionnel et nous recommandons fortement, d'augmenter à deux éléments imagiers l'espacement horizontal entre tous les caractères.

Ouvrages de référence:

Bown, H.G., O'Brien, C.D., Sawchuk, W., et Storey, J.R. Picture Description Instructions (PDIs) for the Telidon Videotex System. Note technique n° 699-E du CRC, novembre 1979

Roudiez, L.S. Readable/Writable/Visible. Visible Language 12(3), 231-244, 1978.

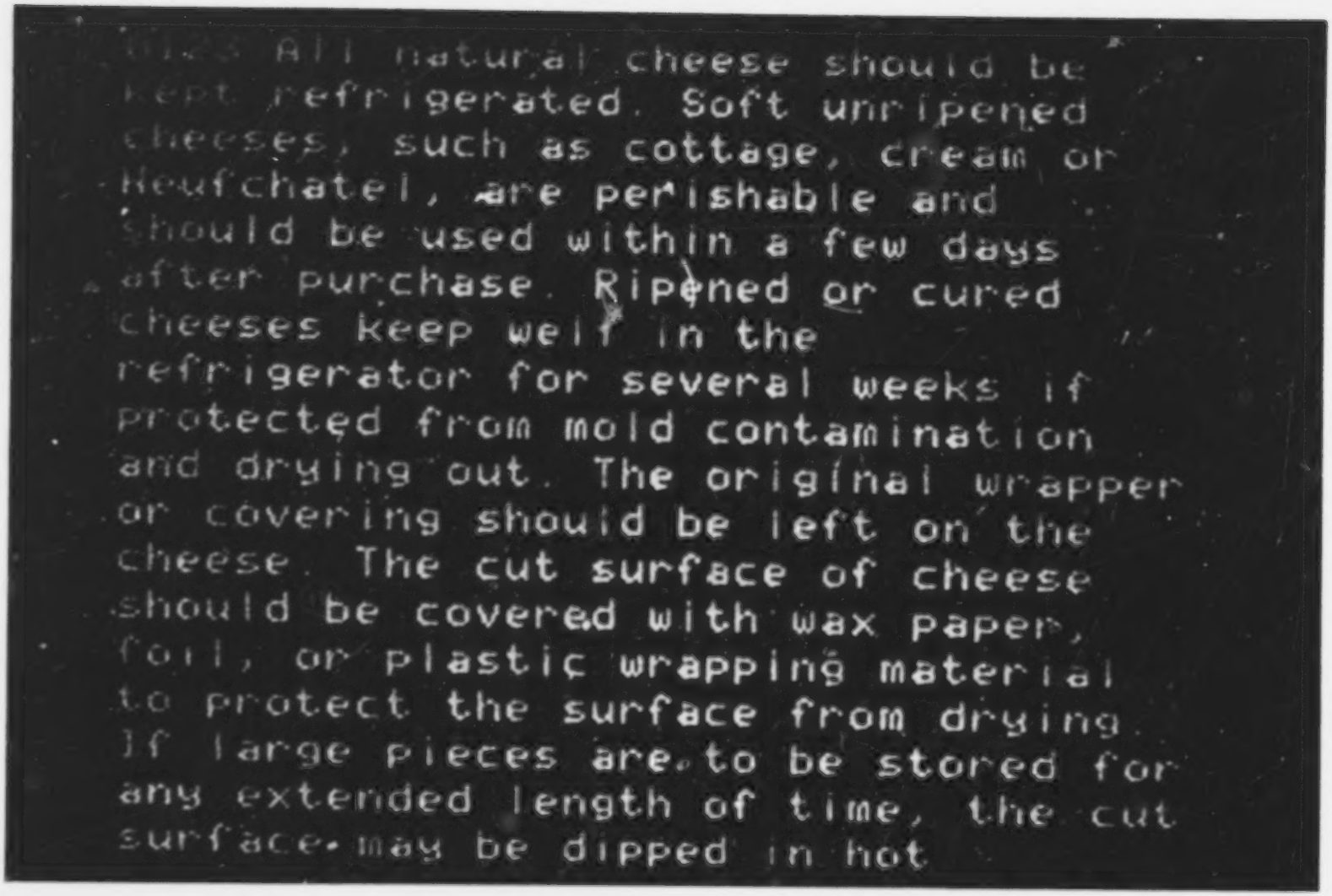
Treurniet, W.C. Spacing of Characters on Television. Proceedings of the Conference on Processing of Visible Language II, New York: Plenum Press, 1979a.

Treurniet, W.C. Espacement des caractères sur un écran de télévision: Jeu à caractères 7 sur 9. Document interne. Ministère des Communications, Canada, 1979b.

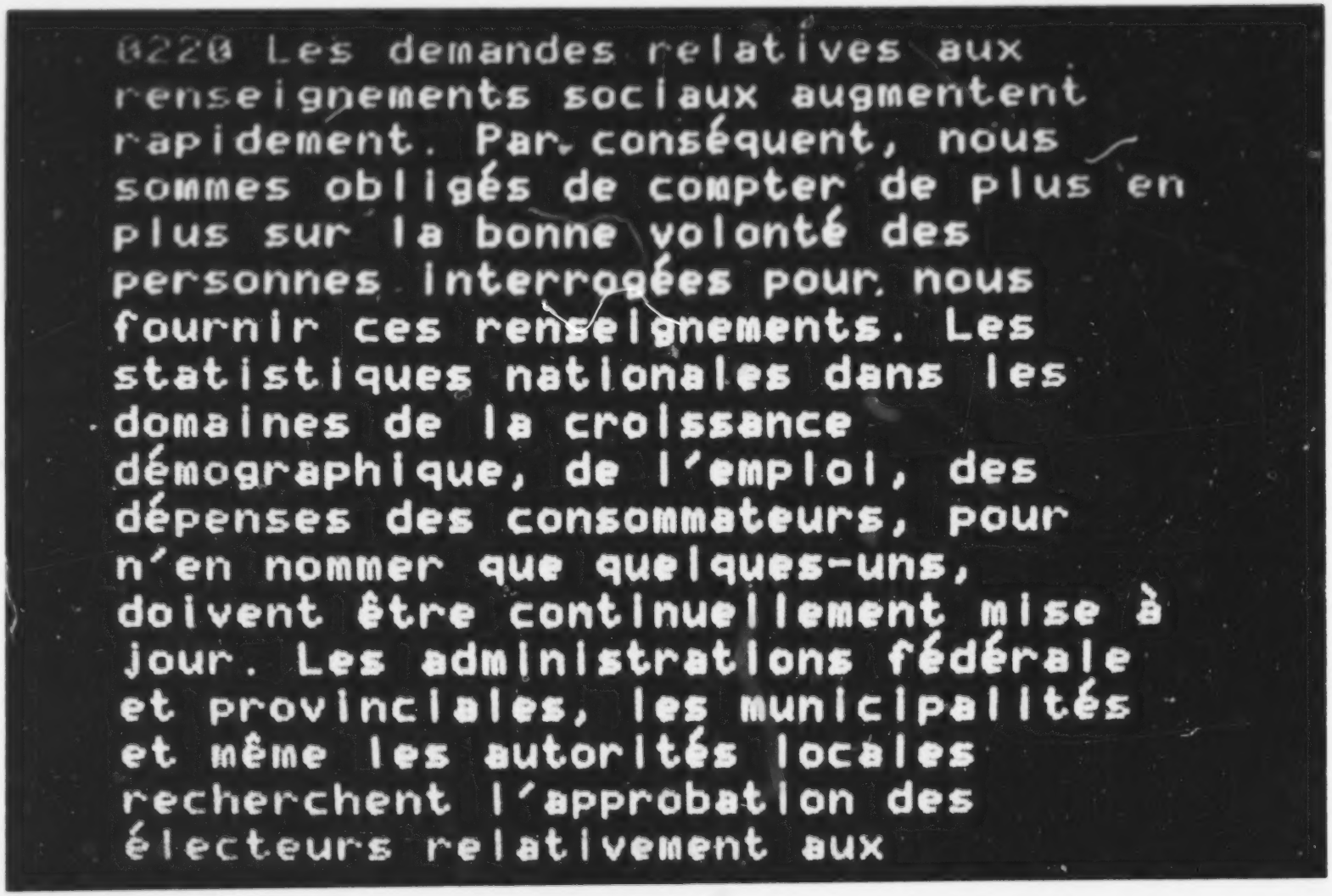
Treurniet, W.C. et Latrémouille, S. Accentuation des caractères. Document interne. Ministère des Communications, Canada, 1979.

ANNEXE A

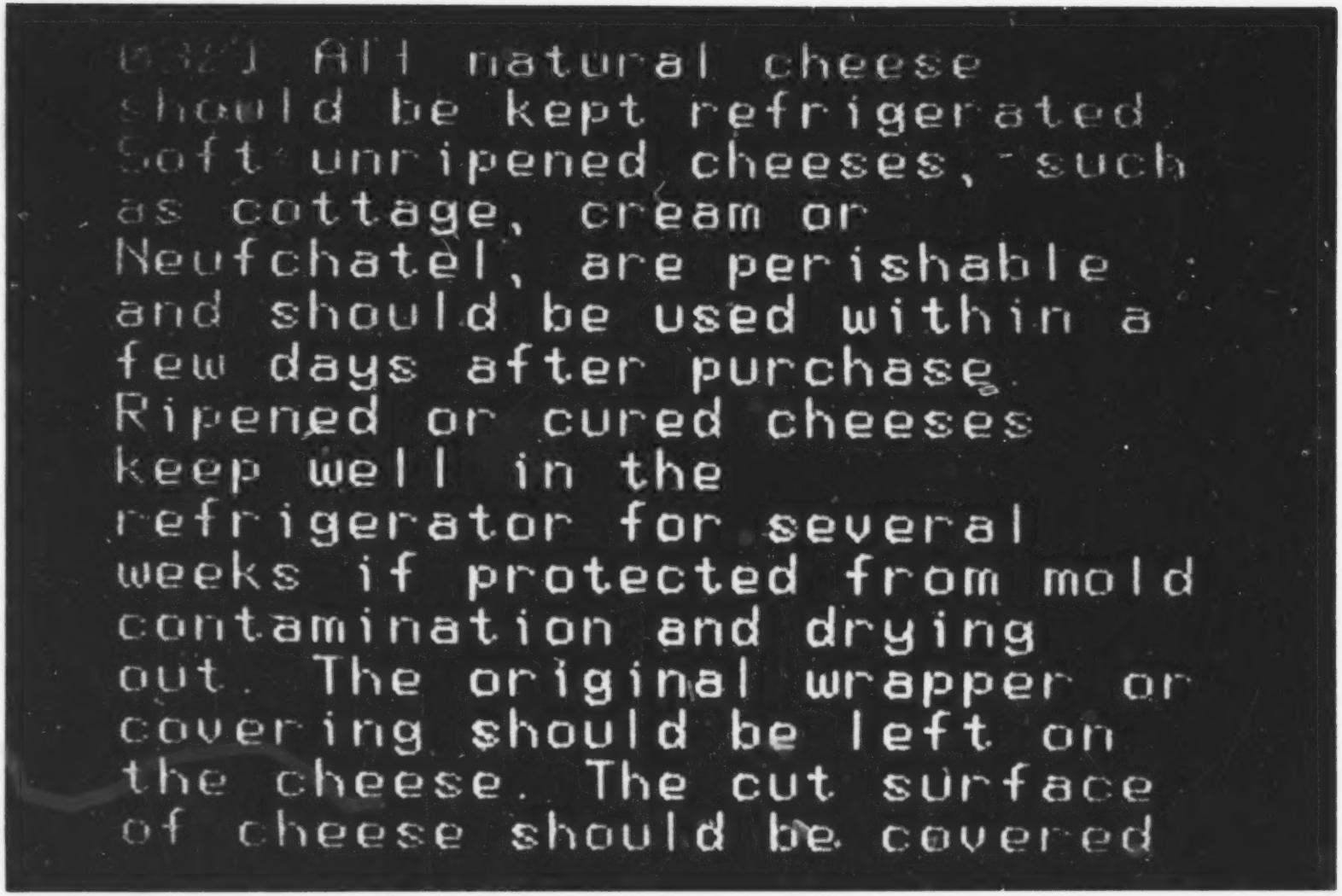
Les figures de la présente annexe montrent l'apparence des textes affichés à l'aide des huit jeux de caractères selon les espacements recommandés. Le nombre indiqué entre parenthèses sur chaque figure indique l'espacement utilisé. Le premier chiffre de ce nombre indique l'espacement horizontal en éléments imagiers, alors que le deuxième chiffre indique l'espacement vertical. Le type de signal d'entrée vidéo est également indiqué par "RGB" (entrée RVB) ou par "CV" (entrée vidéo-composée).



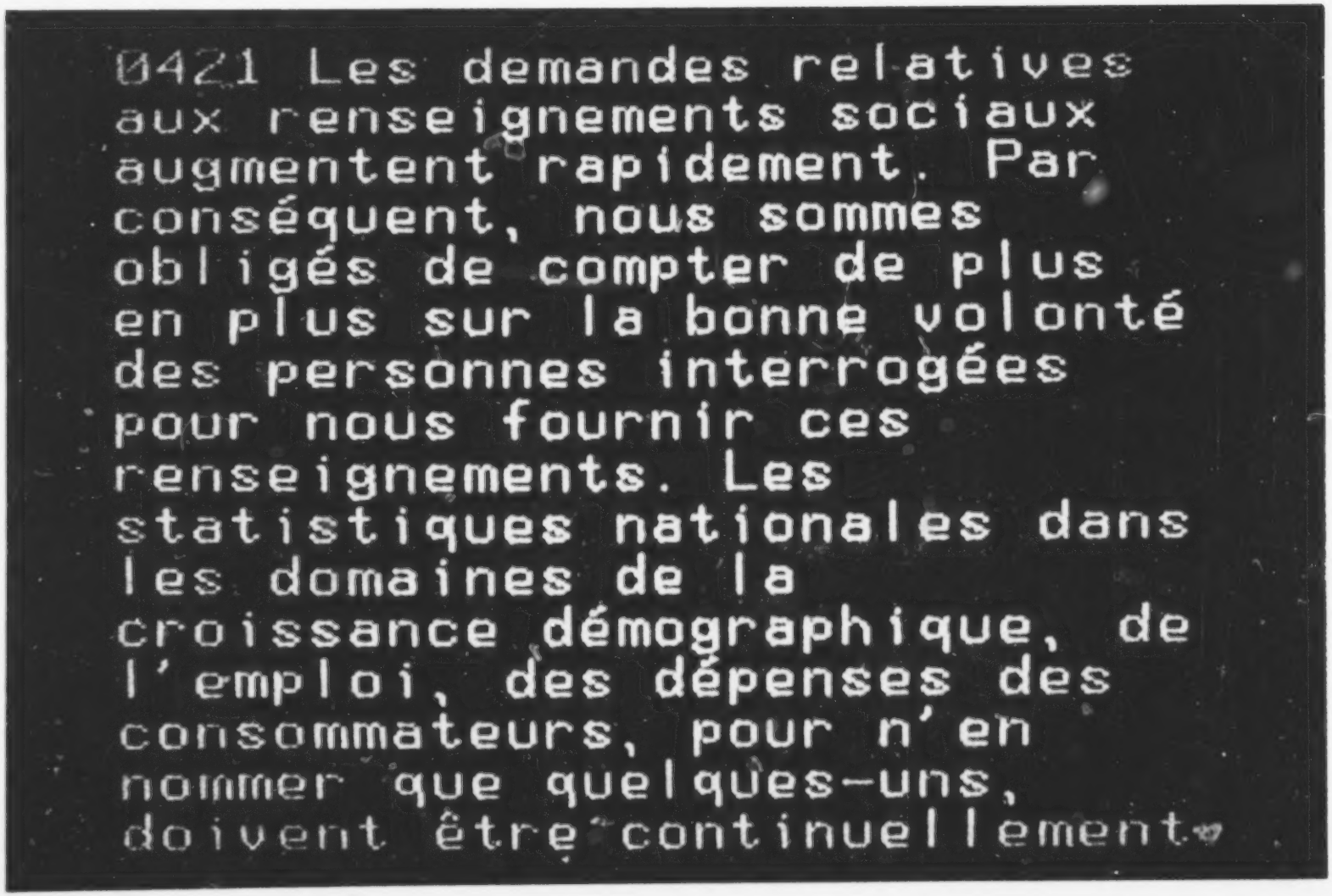
CHSET 1 (23) - RGB



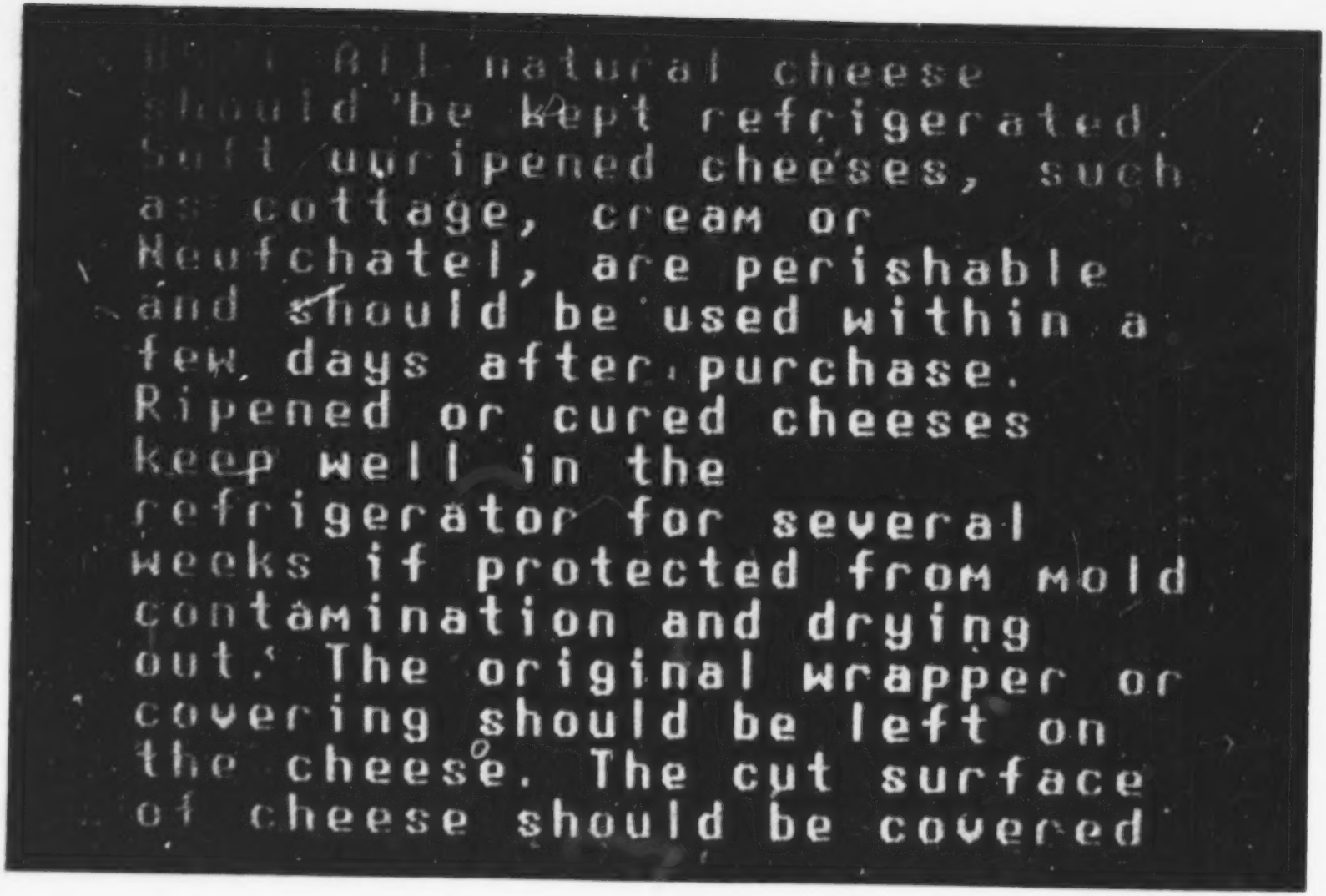
CHSET 2 (20) - RGB



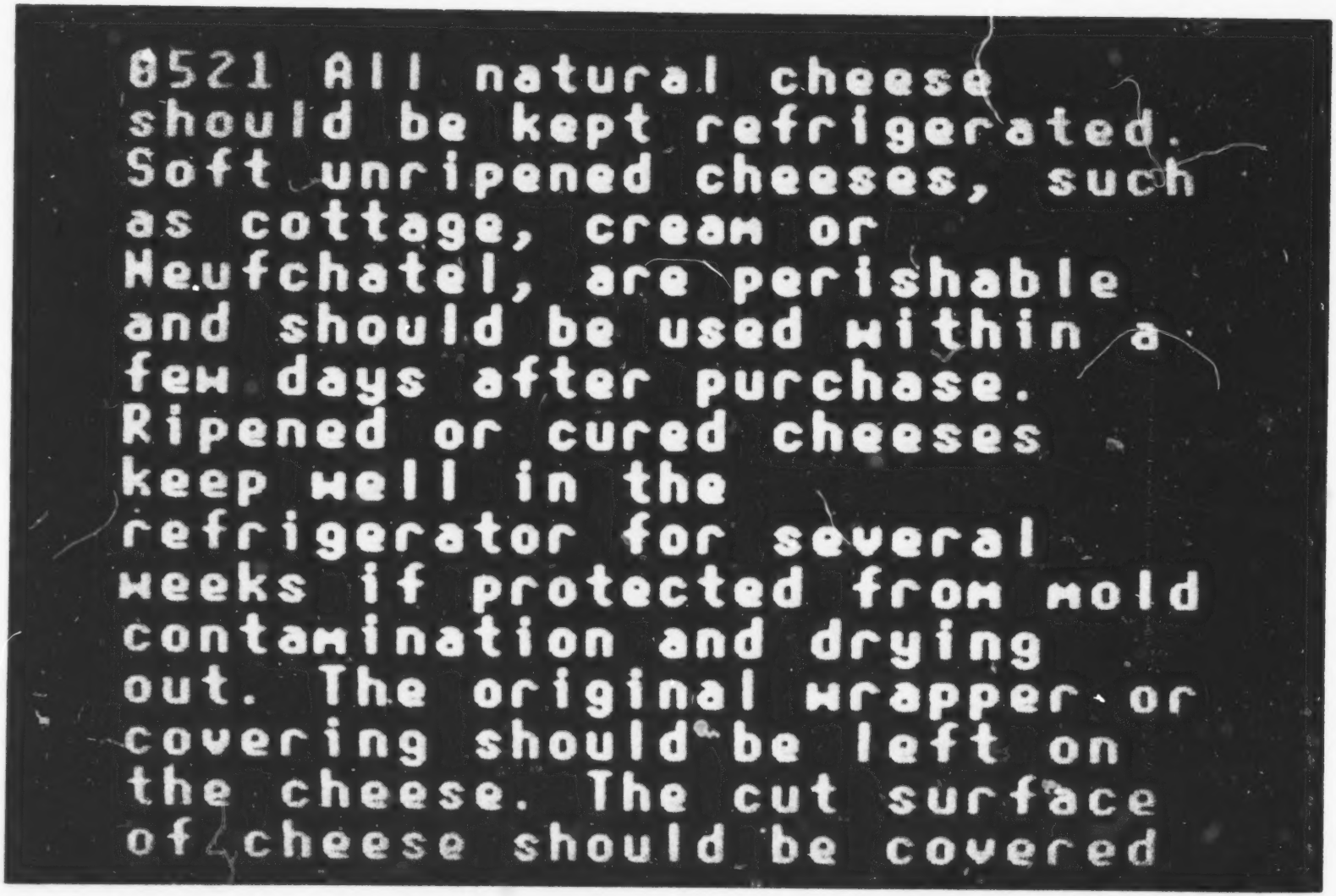
CHSET 3 (21) - RGB

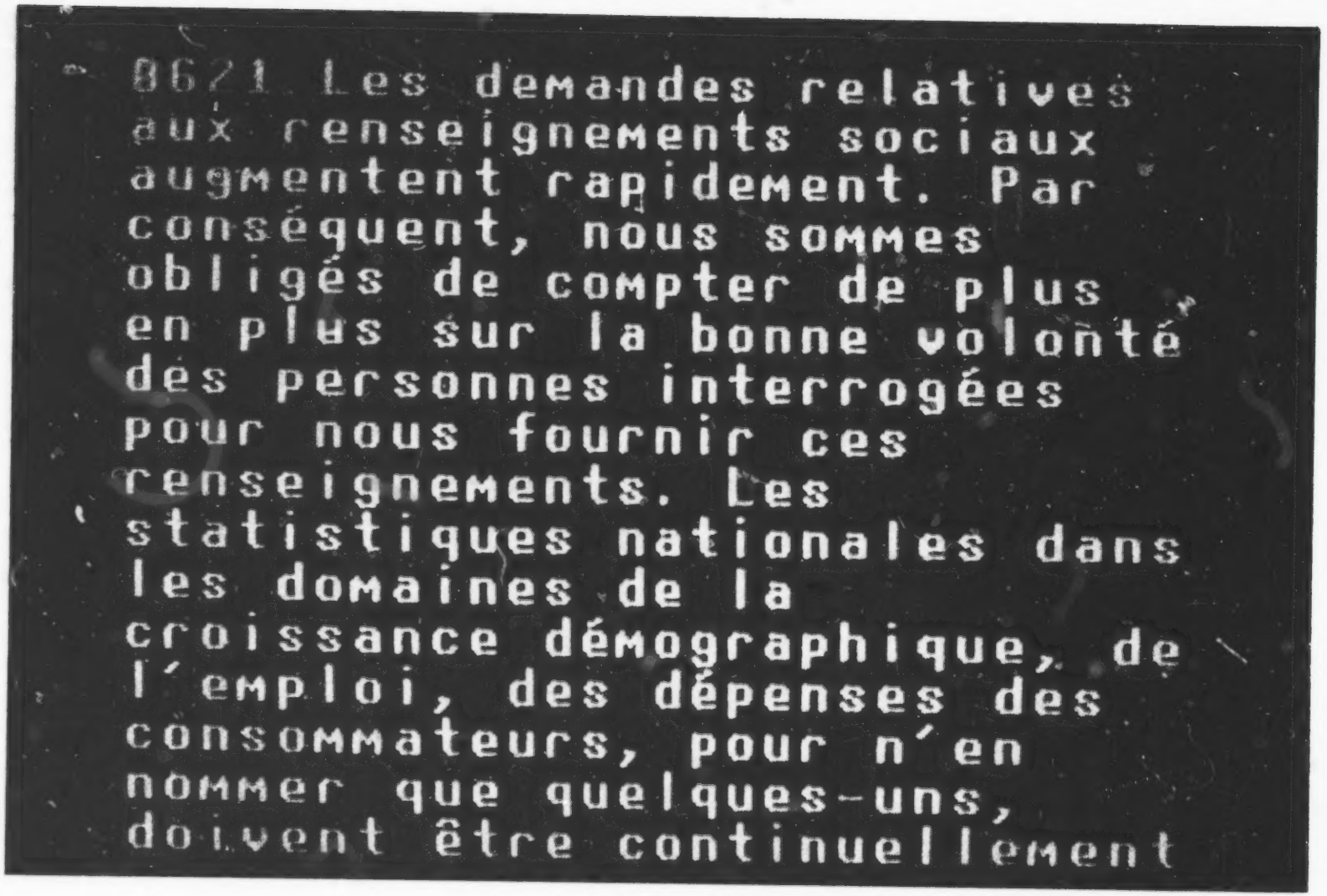


CHSET 4 (21) - RGB

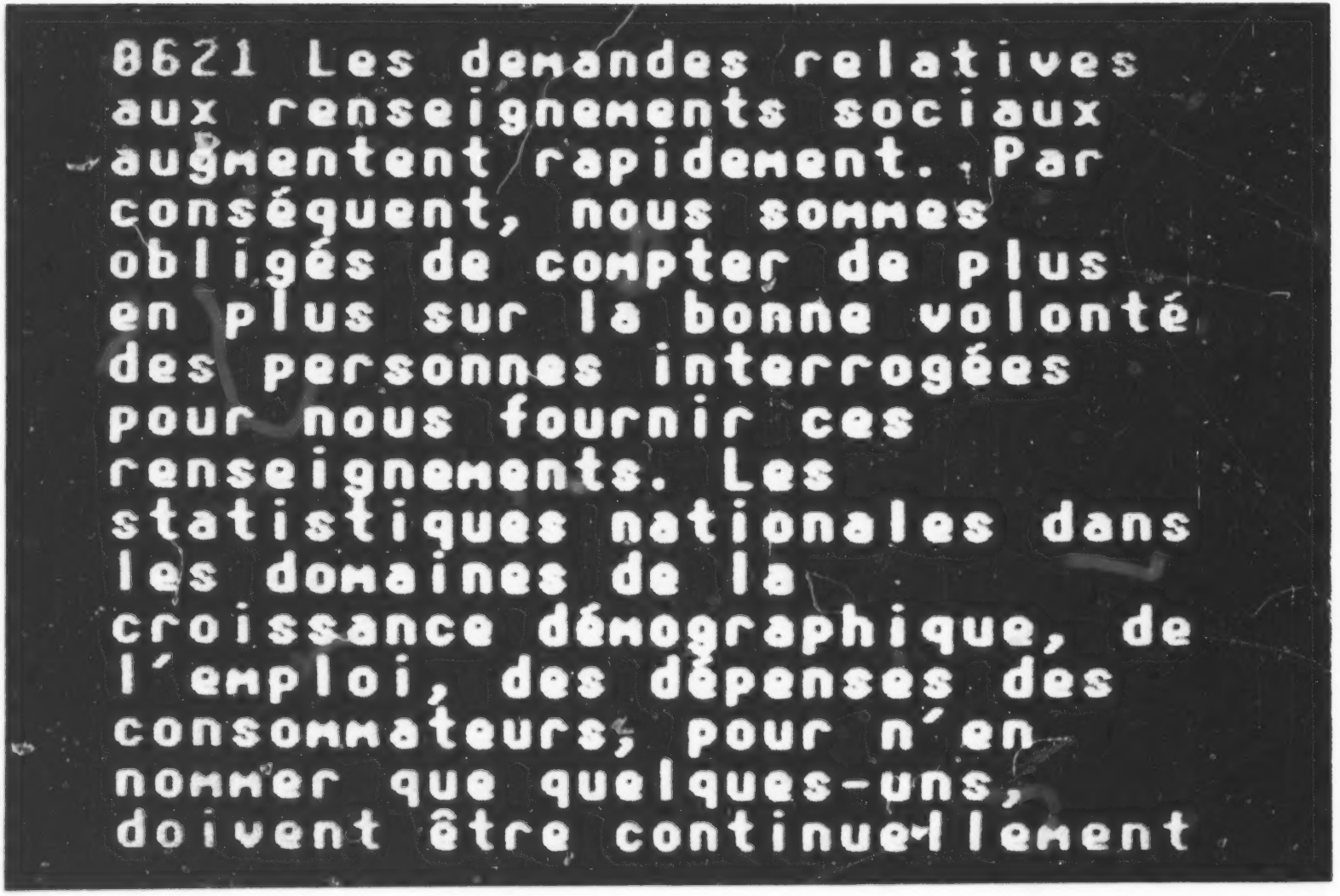


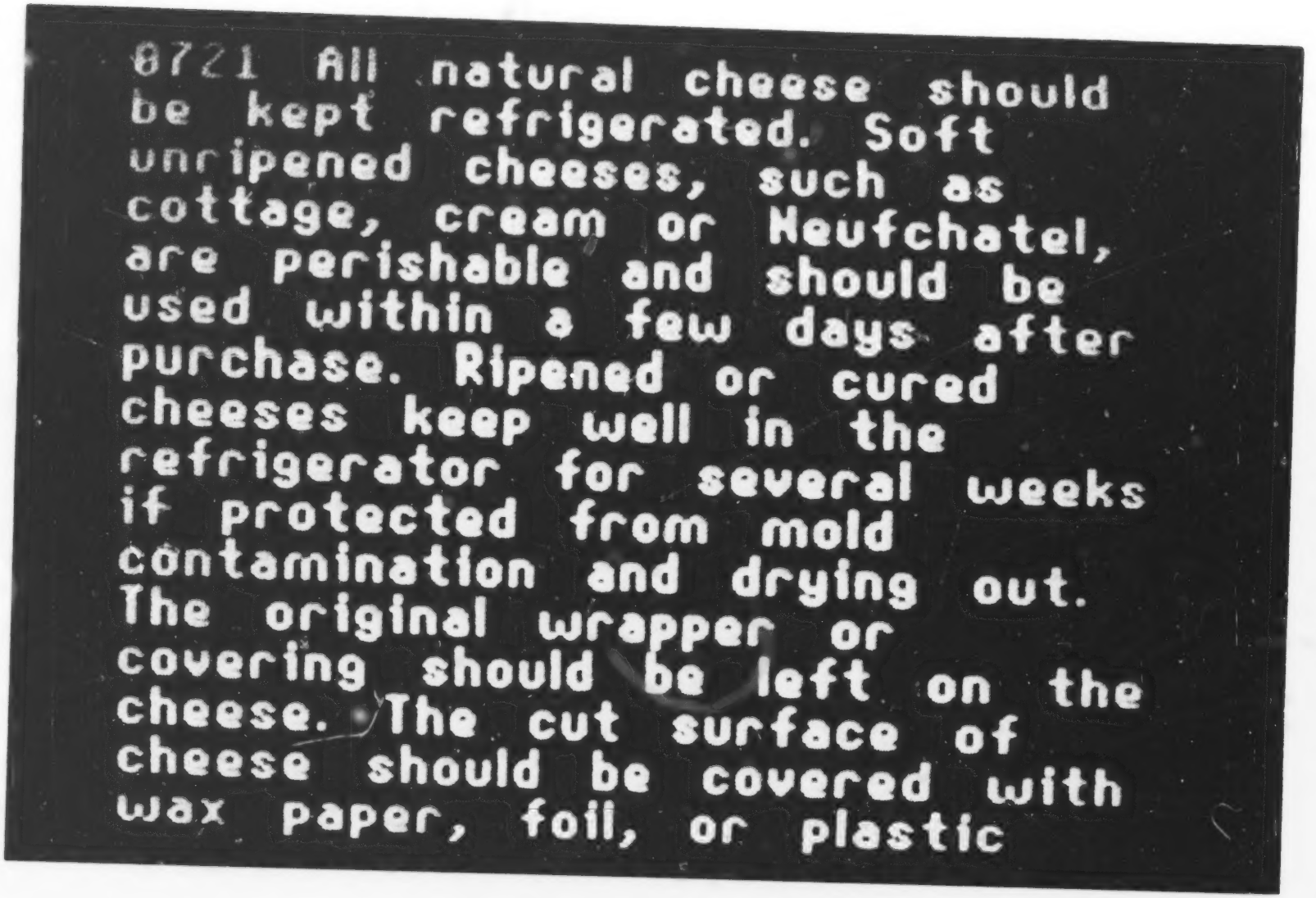
CHSET 5 (21) - RGB


0521 All natural cheese
should be kept refrigerated.
Soft unripened cheeses, such
as cottage, cream or
Neufchatel, are perishable
and should be used within a
few days after purchase.
Ripened or cured cheeses
keep well in the
refrigerator for several
weeks if protected from mold
contamination and drying
out. The original wrapper or
covering should be left on
the cheese. The cut surface
of cheese should be covered

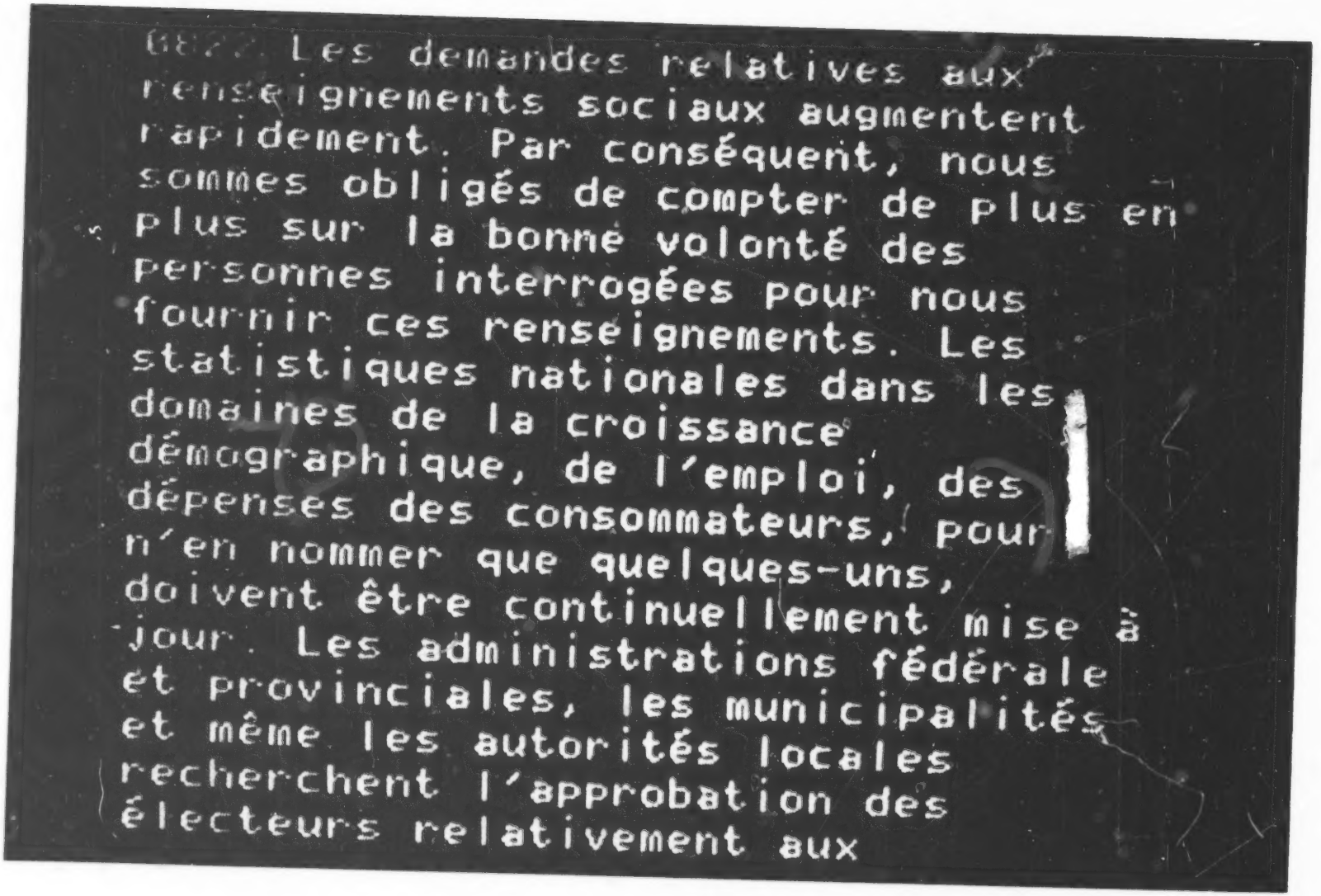


CHSET 6 (21) - RGB


0621 Les demandes relatives
aux renseignements sociaux
augmentent rapidement. Par
conséquent, nous sommes
obligés de compter de plus
en plus sur la bonne volonté
des personnes interrogées
pour nous fournir ces
renseignements. Les
statistiques nationales dans
les domaines de la
croissance démographique, de
l'emploi, des dépenses des
consommateurs; pour n'en
nommer que quelques-uns,
doivent être continuellement


8721 All natural cheese should be kept refrigerated. Soft unripened cheeses, such as cottage, cream or Neufchatel, are perishable and should be used within a few days after purchase. Ripened or cured cheeses keep well in the refrigerator for several weeks if protected from mold contamination and drying out. The original wrapper or covering should be left on the cheese. The cut surface of cheese should be covered with wax paper, foil, or plastic

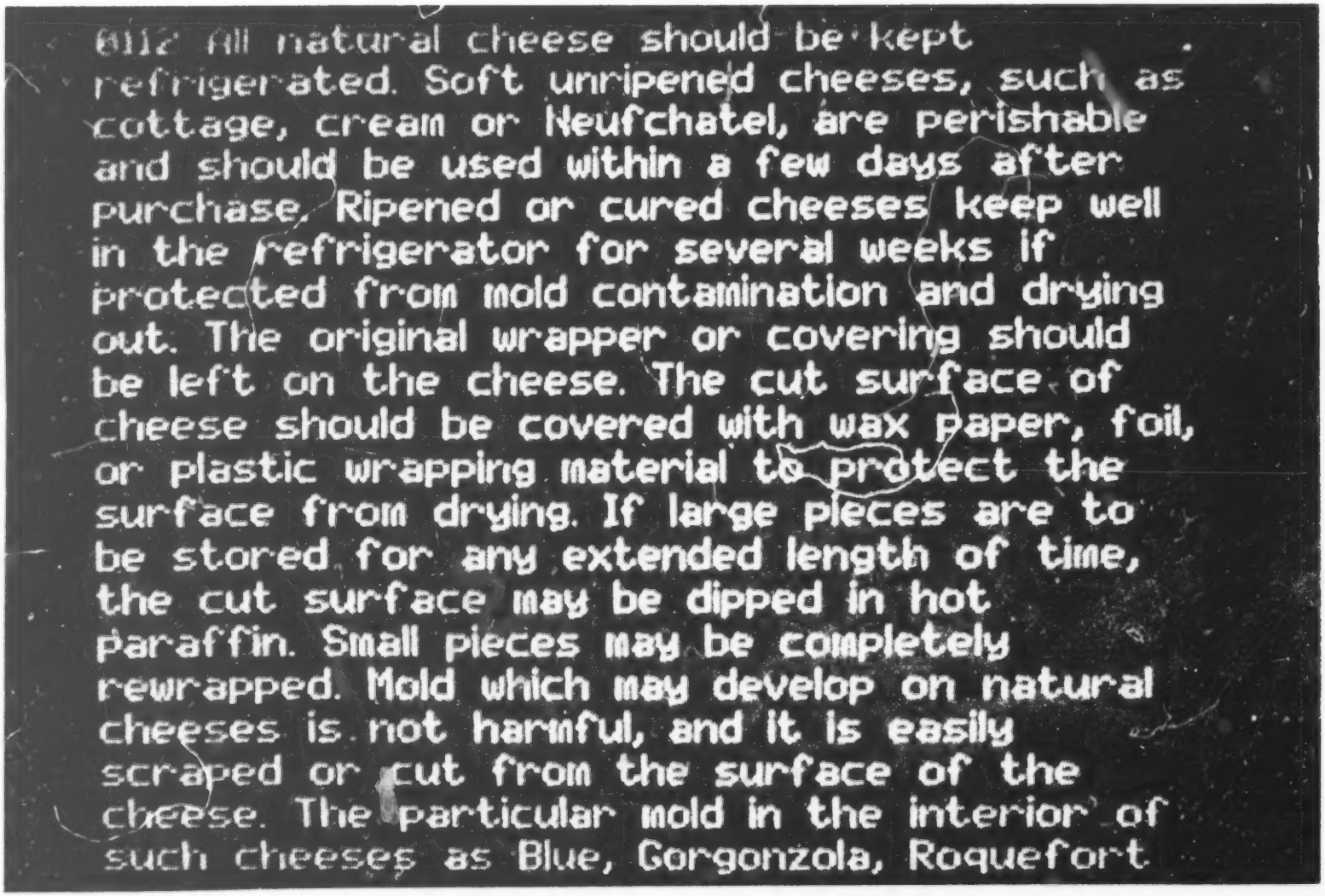
CHSET 7 (21) - CV



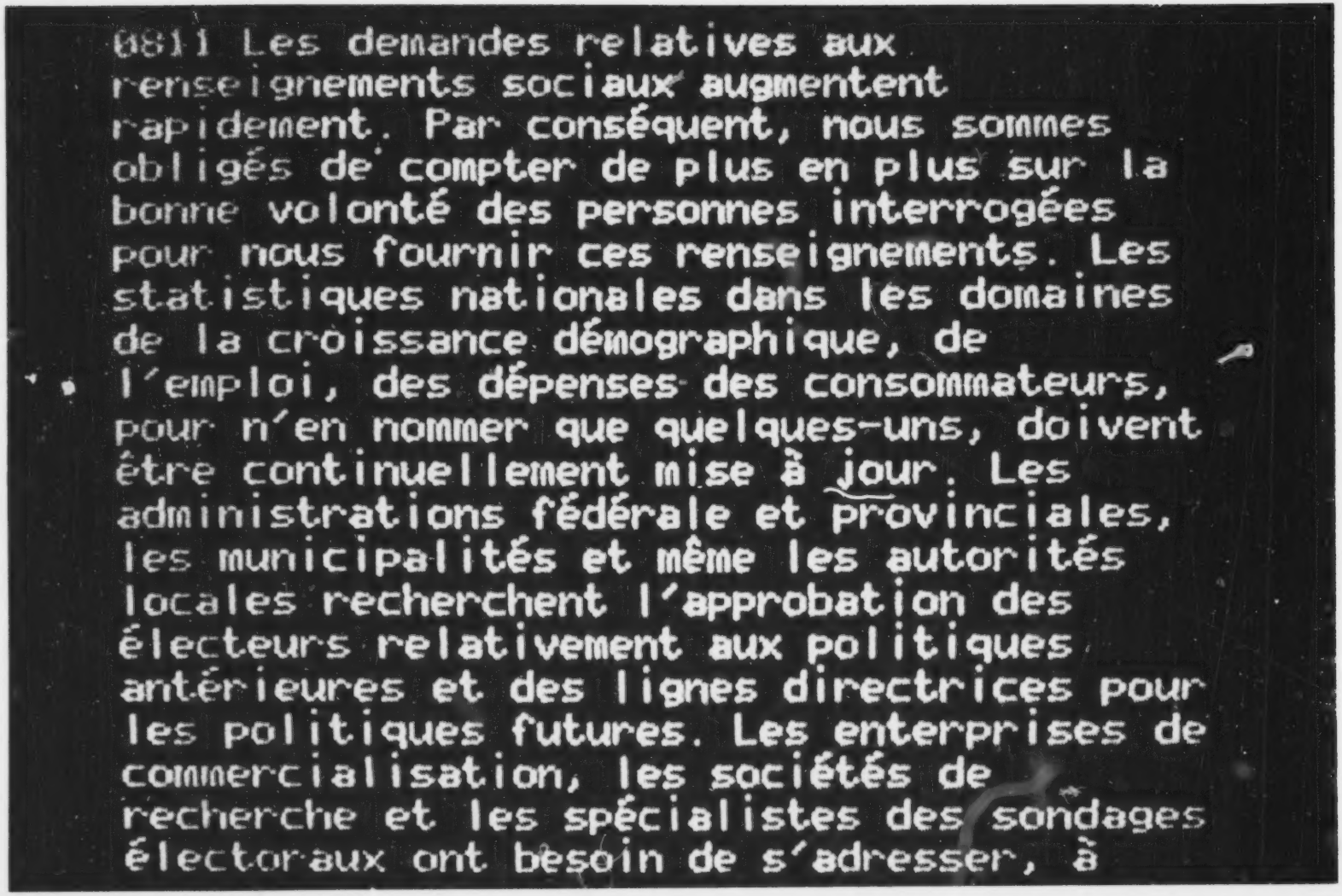
CHSET 8 (22) - RGB

ANNEXE B

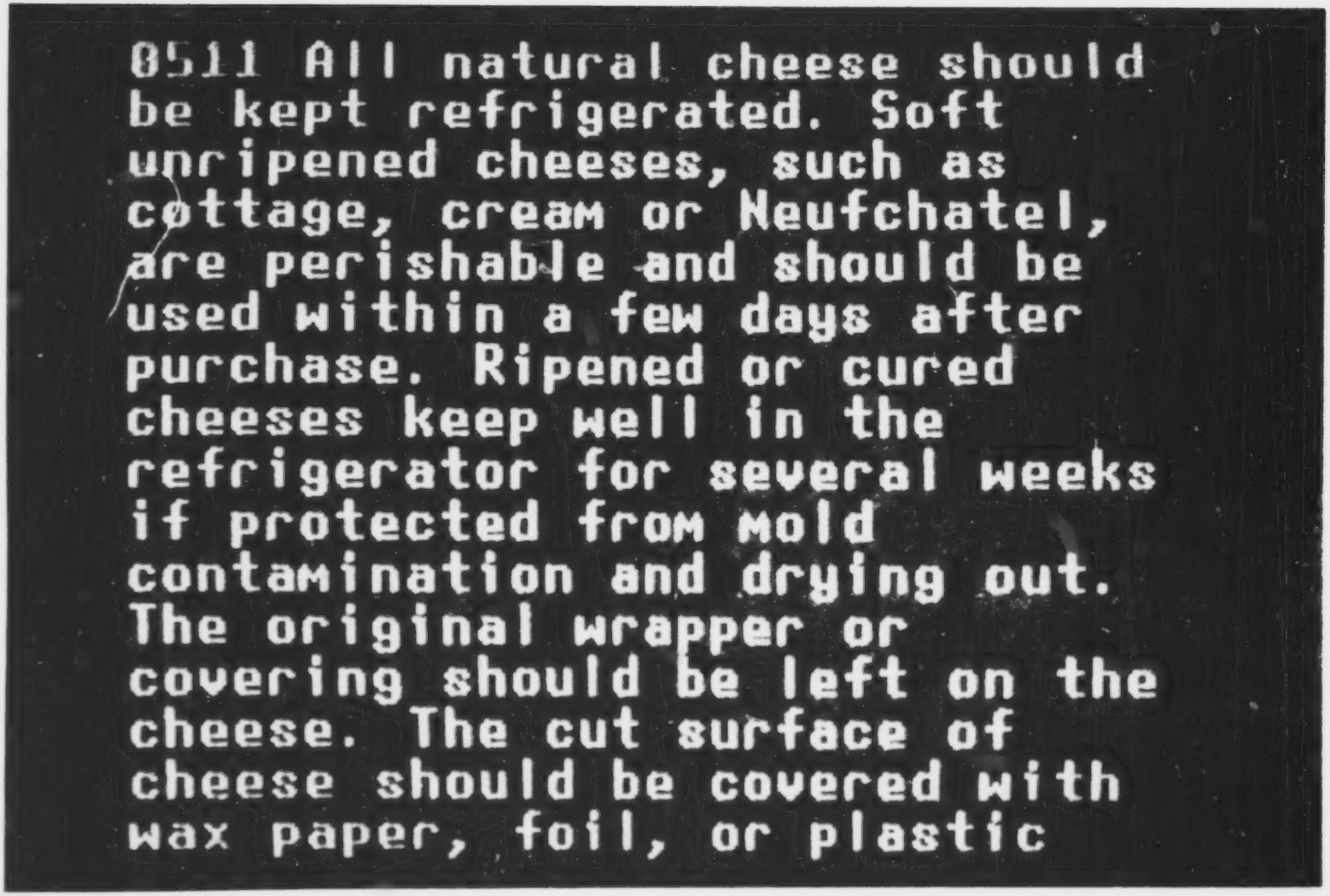
Les figures de la présente annexe montrent l'apparence des textes affichés à l'aide des jeux de caractères et des espacements recommandés pour les expériences-pilotes. Les espacements permettent d'afficher 20 lignes de 40 caractères ou 16 lignes de 32 caractères. Les explications sur la présentation des figures sont les mêmes qu'à l'Annexe A.



CHSET 1 (12) - RGB

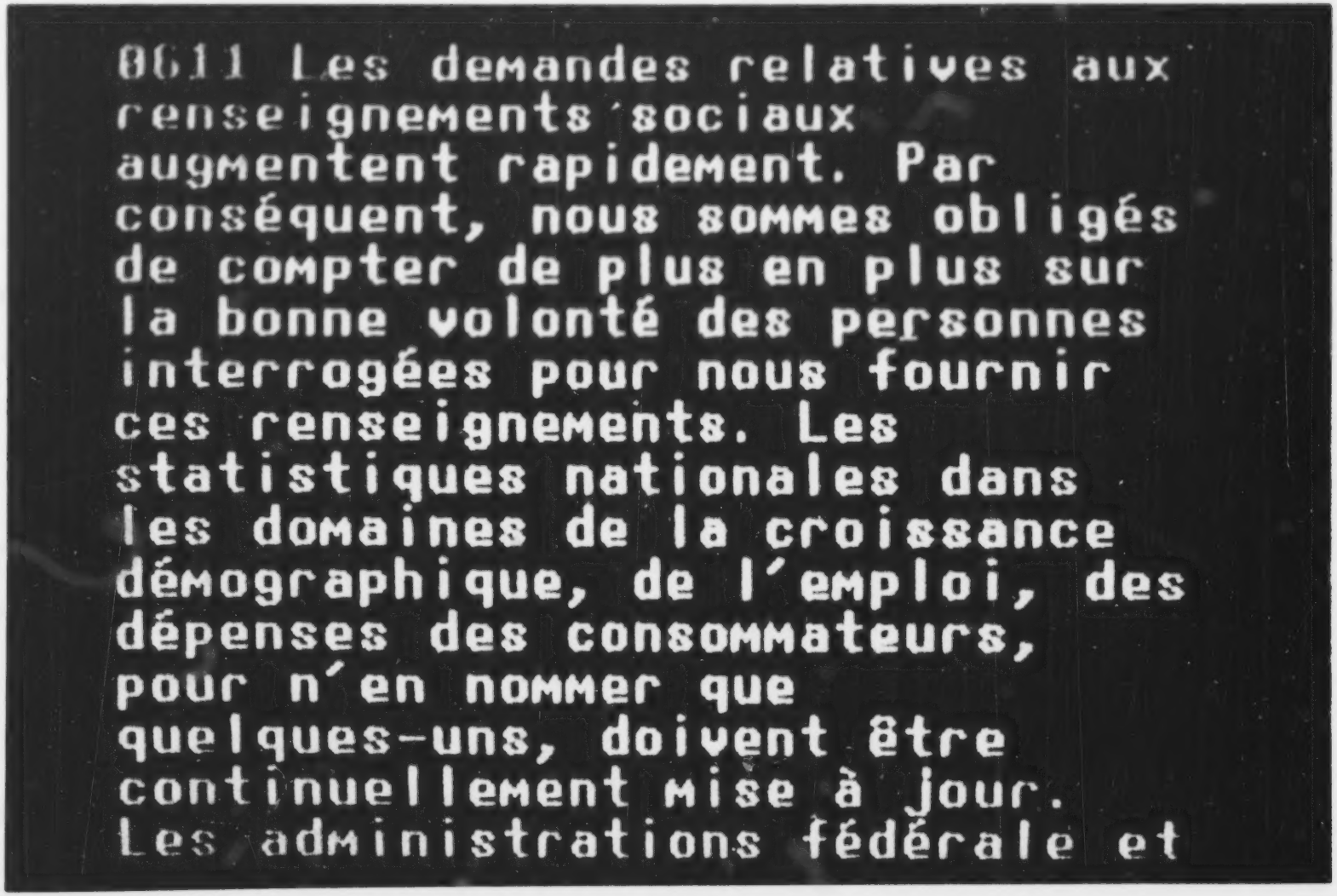


CHSET 8 (11) - RGB


0511 All natural cheese should
be kept refrigerated. Soft
unripened cheeses, such as
cottage, cream or Neufchatel,
are perishable and should be
used within a few days after
purchase. Ripened or cured
cheeses keep well in the
refrigerator for several weeks
if protected from mold
contamination and drying out.
The original wrapper or
covering should be left on the
cheese. The cut surface of
cheese should be covered with
wax paper, foil, or plastic

0511 All natural cheese should be kept refrigerated. Soft unripened cheeses, such as cottage, cream or Neufchatel, are perishable and should be used within a few days after purchase. Ripened or cured cheeses keep well in the refrigerator for several weeks if protected from mold contamination and drying out. The original wrapper or covering should be left on the cheese. The cut surface of cheese should be covered with wax paper, foil, or plastic

CHSET 5 (11) - CV

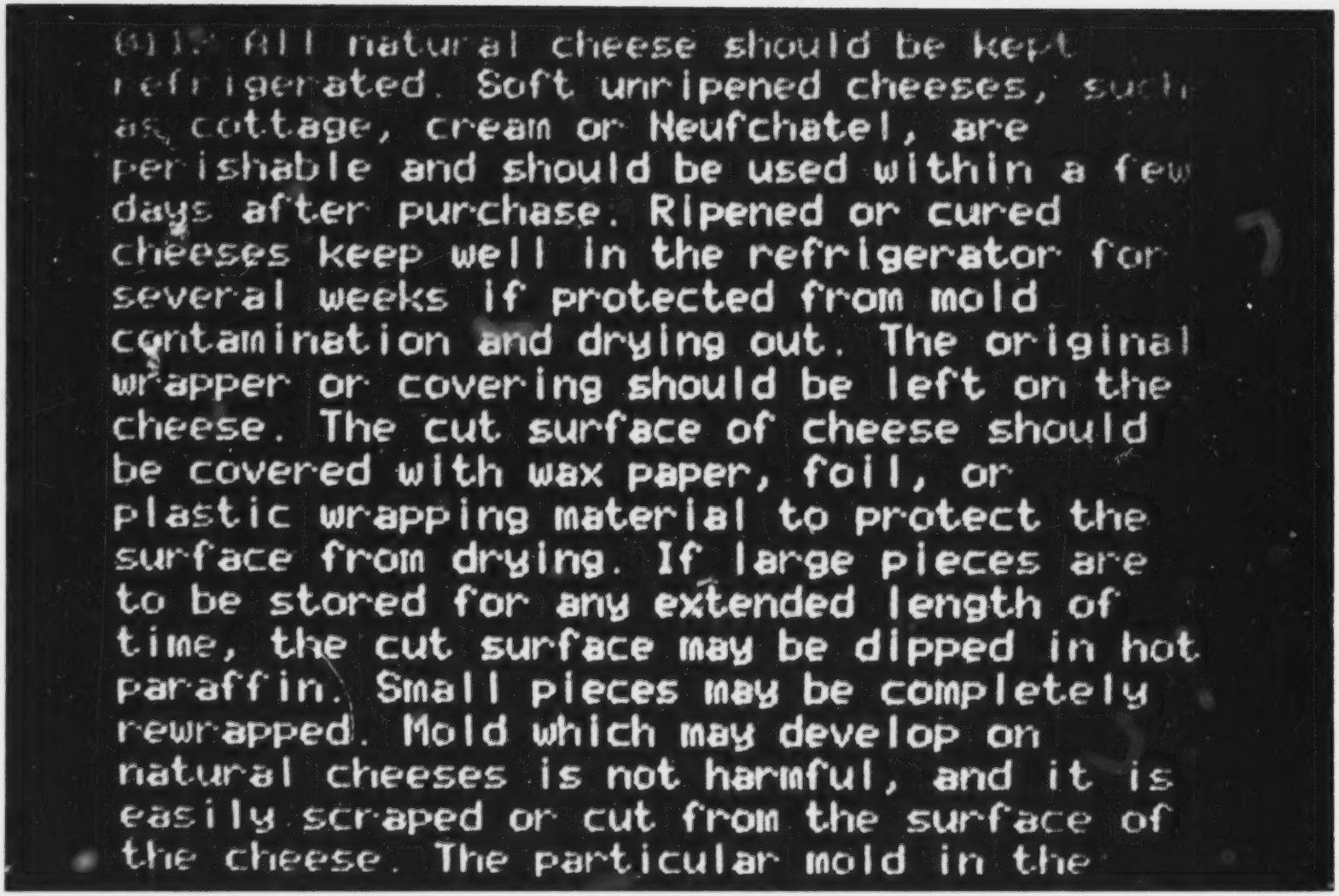


CHSET 6 (11) - RGB

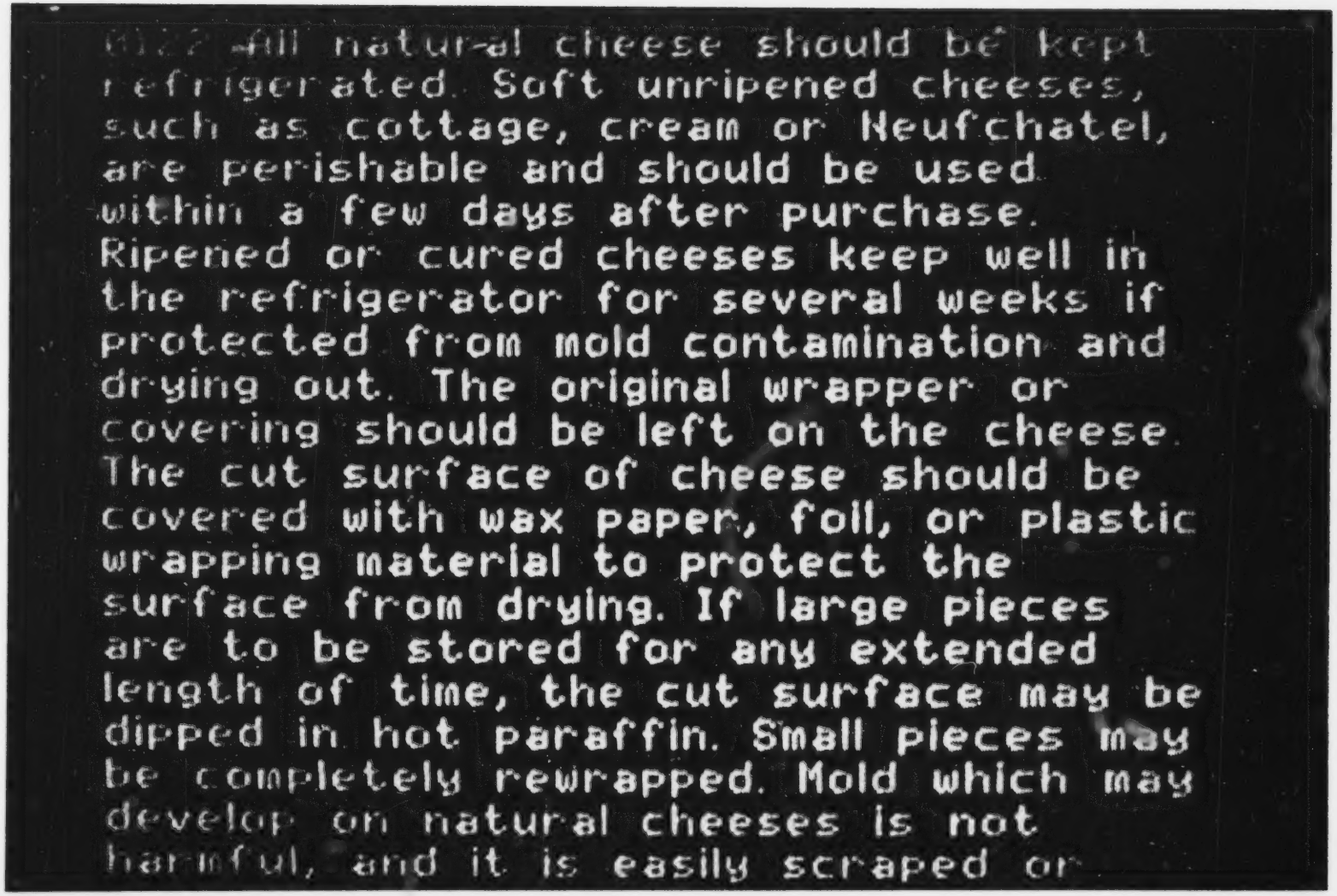
0611 Les demandes relatives aux renseignements sociaux augmentent rapidement. Par conséquent, nous sommes obligés de compter de plus en plus sur la bonne volonté des personnes interrogées pour nous fournir ces renseignements. Les statistiques nationales dans les domaines de la croissance démographique, de l'emploi, des dépenses des consommateurs, pour n'en nommer que quelques-uns, doivent être continuellement mise à jour. Les administrations fédérale et

ANNEXE C

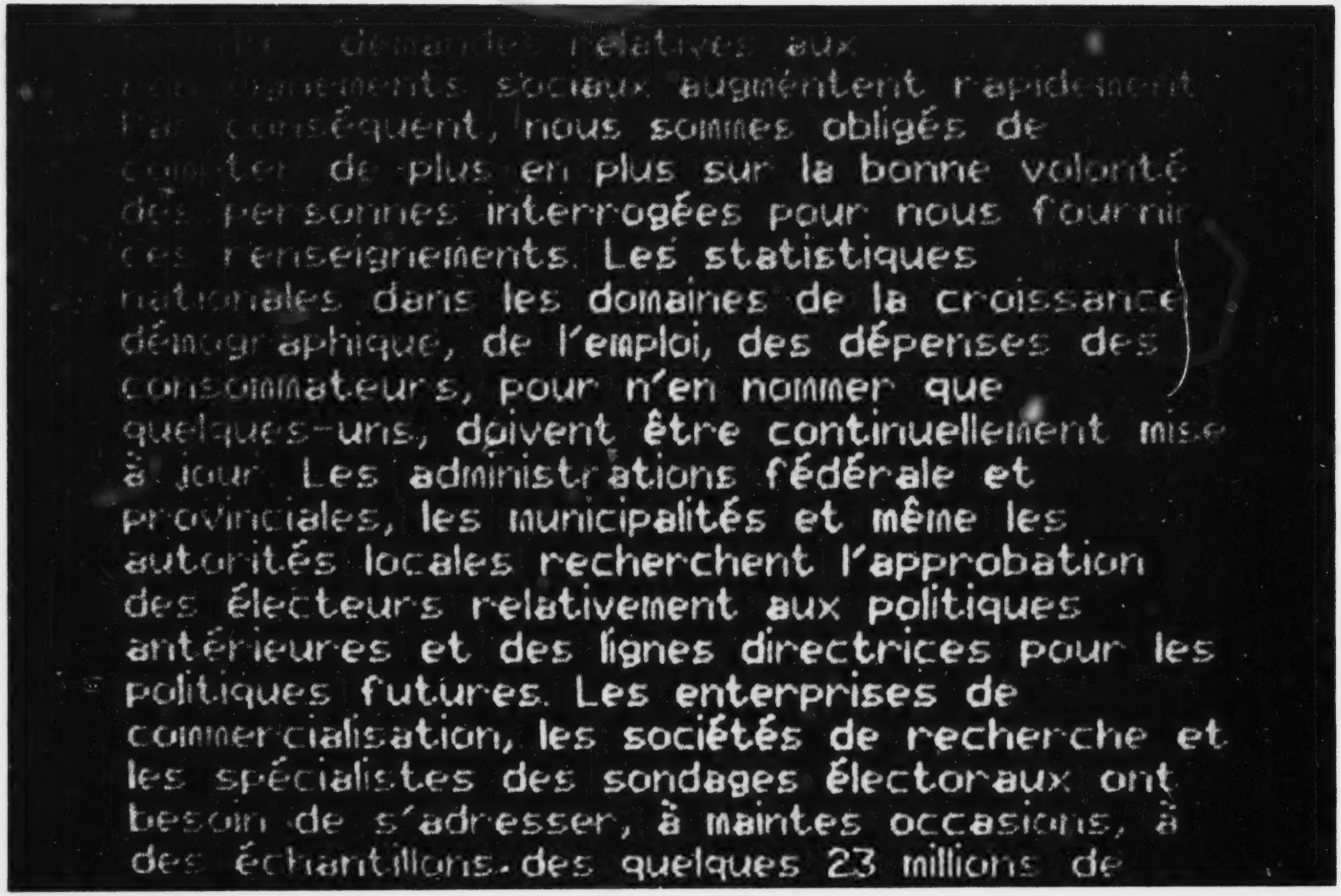
Les figures de la présente annexe montrent l'effet de l'espacement proportionnel sur la lisibilité des textes. Elles montrent également l'effet salutaire obtenu en augmentant d'un élément imagier l'espacement horizontal. Dans ce cas également, les explications sur la présentation des figures sont les mêmes qu'à l'Annexe A.



CHSET 1 (12) - RGB



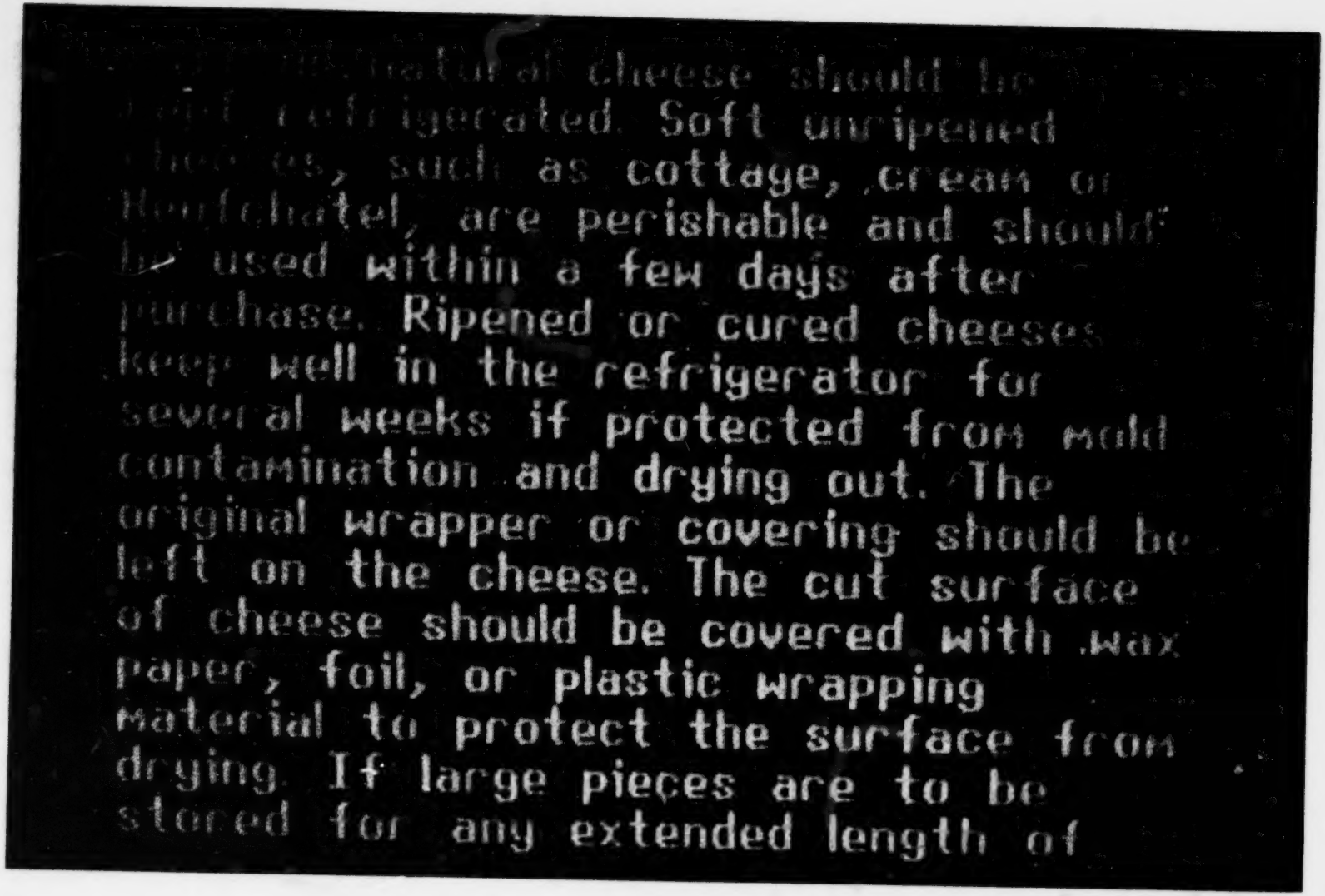
CHSET 1 (22) - RGB



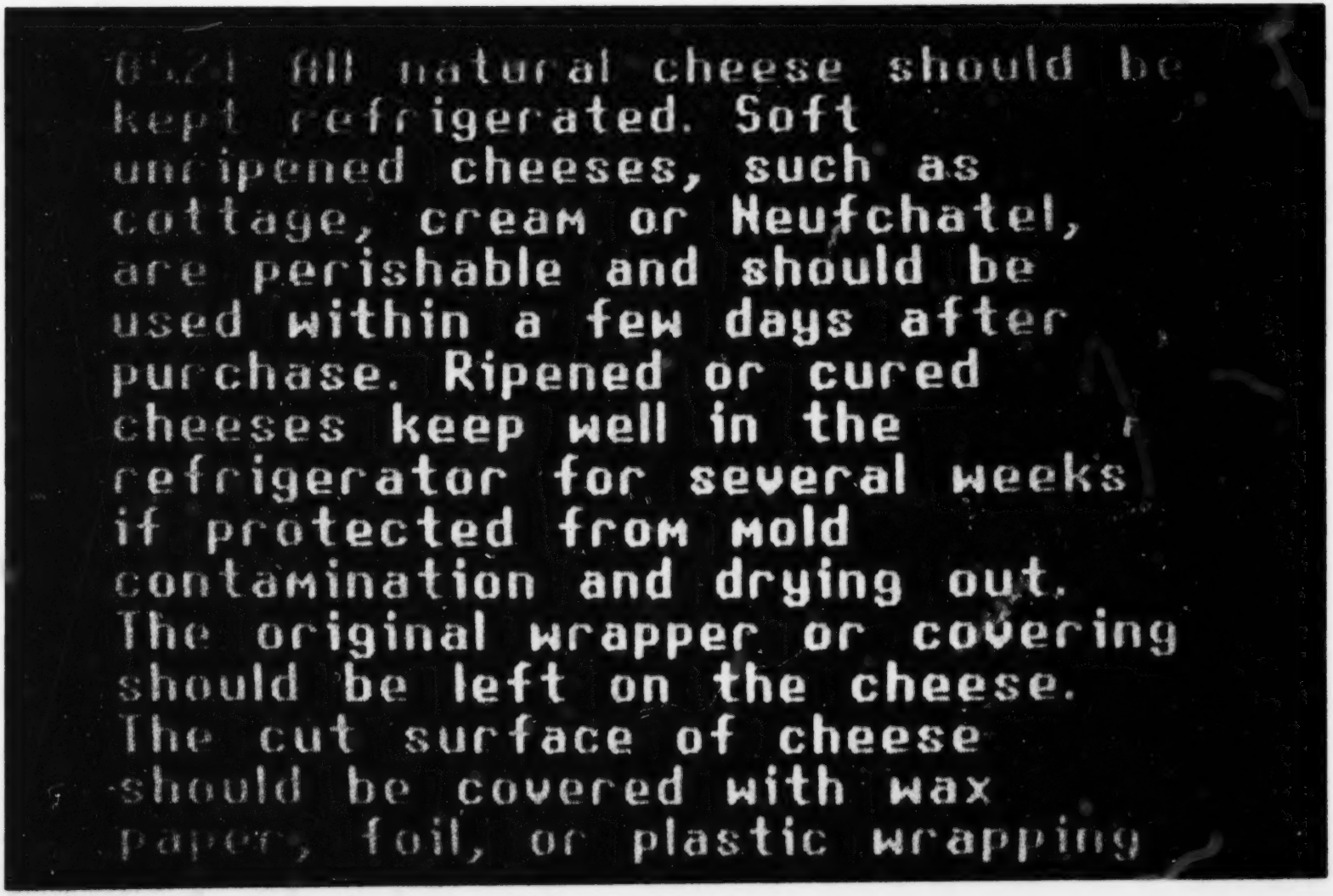
CHSET 8 (11) - RGB

8821 Les demandes relatives aux renseignements sociaux augmentent rapidement. Par conséquent, nous sommes obligés de compter de plus en plus sur la bonne volonté des personnes interrogées pour nous fournir ces renseignements. Les statistiques nationales dans les domaines de la croissance démographique, de l'emploi, des dépenses des consommateurs, pour n'en nommer que quelques-uns, doivent être continuellement mise à jour. Les administrations fédérale et provinciales, les municipalités et même les autorités locales recherchent l'approbation des électeurs relativement aux politiques antérieures et des lignes directrices pour les politiques futures. Les enterprises de

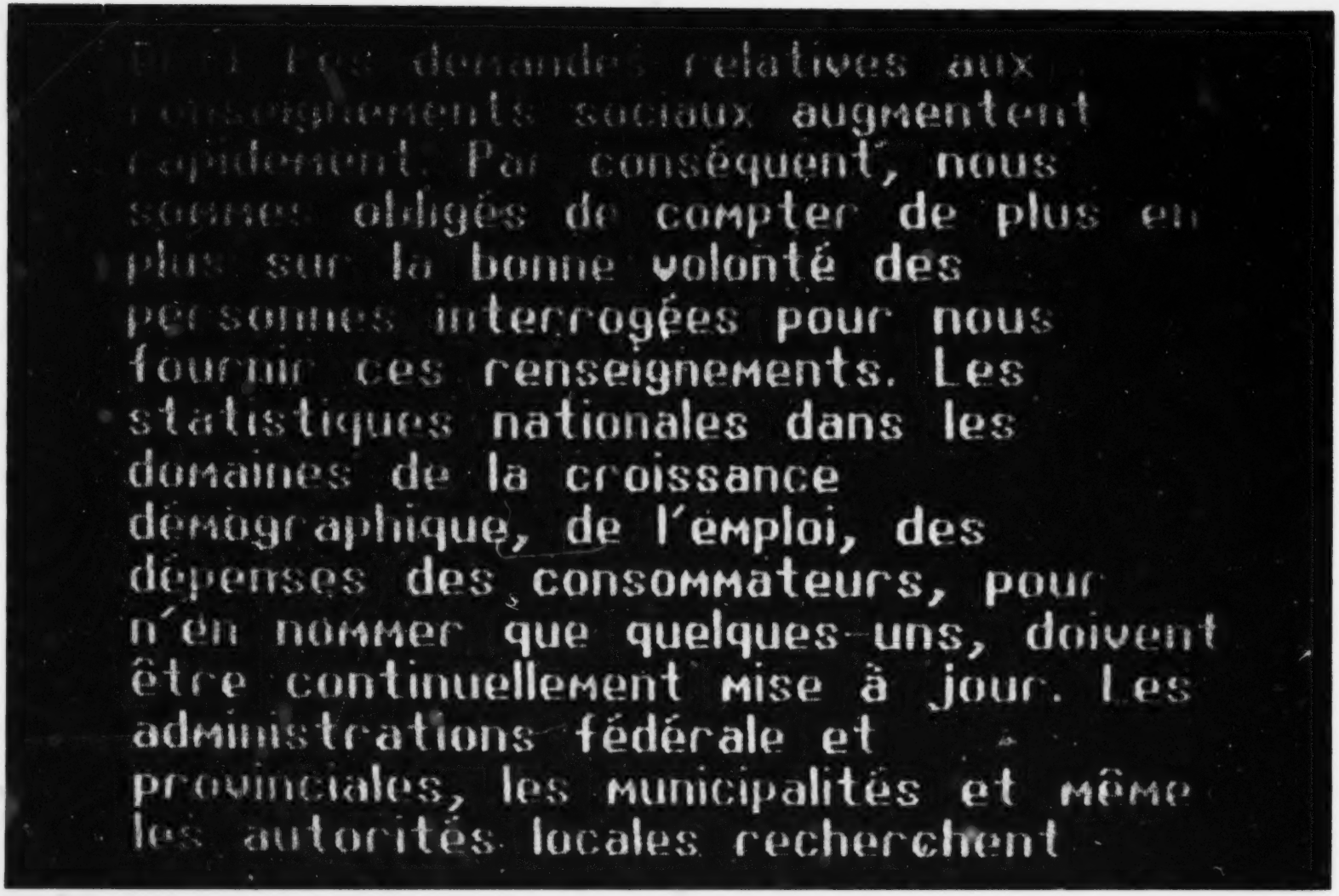
CHSET 8 (21) - RGB



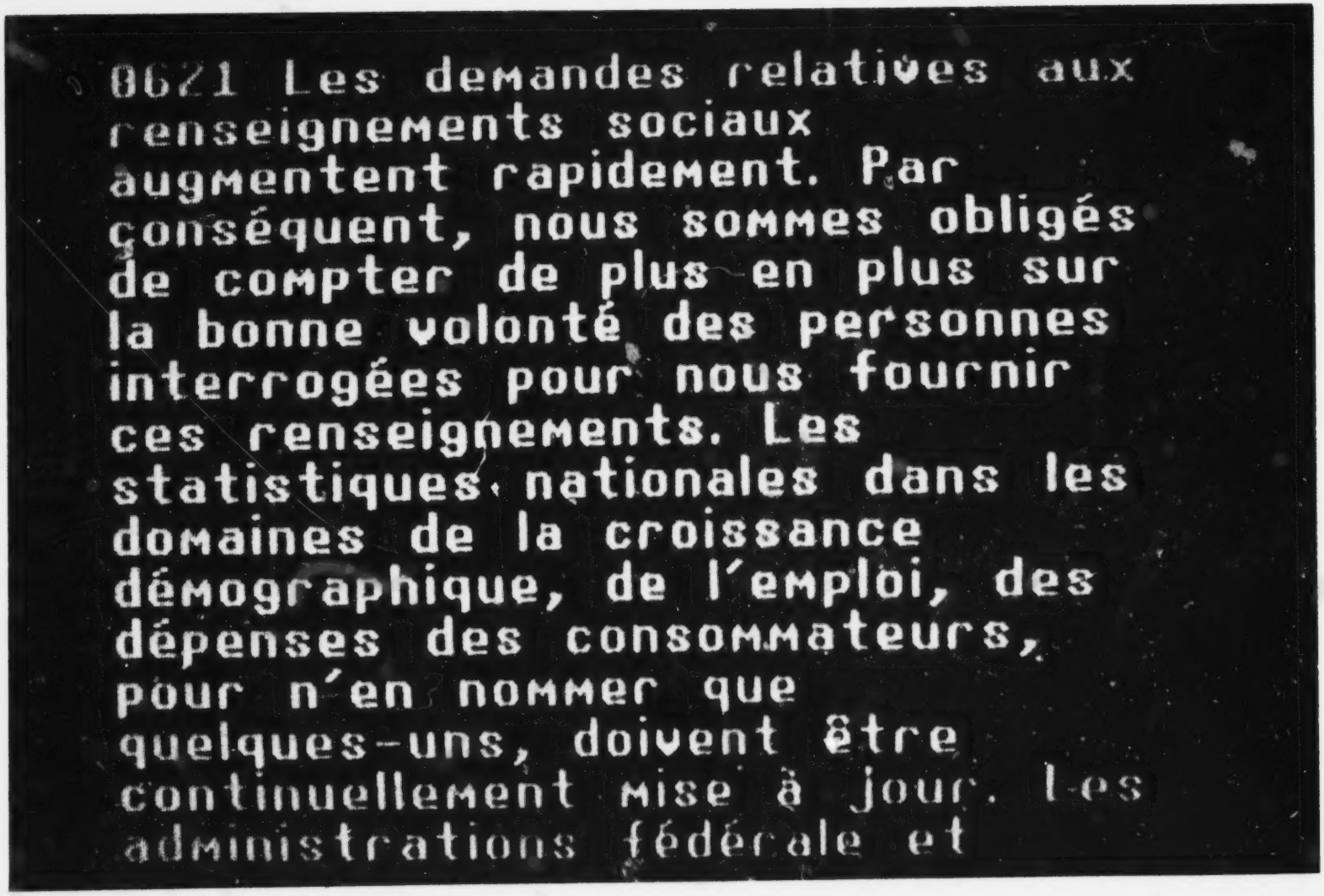
CHSET 5 (11) - RGB



CHSET 5 (21) ~ RGB



CHSET 6 (11) - RGB



CHSET 6 (21) - RGB

0621 Les demandes relatives aux renseignements sociaux augmentent rapidement. Par conséquent, nous sommes obligés de compter de plus en plus sur la bonne volonté des personnes interrogées pour nous fournir ces renseignements. Les statistiques nationales dans les domaines de la croissance démographique, de l'emploi, des dépenses des consommateurs, pour n'en nommer que quelques-uns, doivent être continuellement mise à jour. Les administrations fédérale et

Espacement des caractères sur un écran de télévision
Jeu de caractères 7 sur 9

William C. Treurniet

Centre de recherches sur les communications

Ottawa (Ontario)

Août 1979

Résumé de l'étude

Une expérience précédente menée avec un jeu de caractères 5 sur 7 a établi qu'une page d'affichage Télidon ne devrait pas contenir plus de 20 lignes et plus de 40 caractères la ligne. L'expérience a aussi montré que les caractères descendants devraient dépasser la ligne d'au moins un élément imagier. Le présent document expose les résultats d'une expérience semblable avec un jeu de caractères 7 sur 9.

Notre expérience avec le jeu de caractères 7 sur 9 a montré que les caractères descendants devraient dépasser la ligne d'au moins deux éléments imagiers, que les caractères d'une ligne devraient être séparés par trois éléments imagiers, et que l'espacement entre les lignes pouvait être réduit à un élément imagier. Ces données font passer la matrice de base des caractères 7 sur 9 à une matrice de 10 sur 12. Etant donné que la surface utilisable d'un écran de télévision est d'environ 216 éléments imagiers sur 288, les résultats obtenus indique qu'on ne devrait pas afficher sur le Télidon plus de 18 lignes de 24 caractères si l'on utilise ce jeu de caractères.

L'espace de vingt-quatre caractères permet d'afficher seulement quatre mots environ sur une ligne. Lorsqu'il n'est pas permis de couper les mots, les mots plus longs peuvent occasionner de grands trous vides à la fin des lignes. L'utilisation de l'"espacement proportionnel" ou du "crénage" semble réduire cet effet, mais nous devons poursuivre notre recherche pour confirmer cette impression. Le crénage consiste à réduire la largeur de la matrice des caractères dans le cas des lettres étroites comme le "i" et le "l". S'il est techniquement impossible de faire du crénage sur le Télidon, il pourra être nécessaire d'accepter l'espacement horizontal de deux éléments imagiers. La matrice 9 sur 12 qui en résulterait permettrait d'aligner 32 caractères sur chaque ligne.

Dans un exposé intitulé "Espacement des caractères sur un écran de télévision", nous avons déjà fait état d'une épreuve de recherche de lettres utilisée pour établir nos recommandations concernant l'espacement des caractères sur un écran de télévision. Dans le cas du jeu de caractères 5 sur 7 utilisé pour cette étude, nous avons conclu que le meilleur espacement horizontal était de deux éléments imagiers et le meilleur espacement entre les lignes de lettres, de trois éléments imagiers. De plus, l'expérience a montré que les caractères descendants devraient dépasser la ligne d'au moins un élément imagier. Le présent exposé donne les résultats d'une expérience similaire à l'aide du jeu de caractères 7 sur 9 actuellement utilisé sur le Télidon.

Pour notre expérience, nous avons utilisé un télécouleur Electrohome, modèle C40 de 19 pouces, modifié pour l'entrée en Rouge, Vert et Bleu (RVB). Cet appareil est de qualité semblable à celle des téléviseurs domestiques et est muni d'éléments rectangulaires de substance luminescente. Les lettres étaient affichées en blanc sur fond neutre.

Notre méthodologie était identique à celle que nous avons décrite dans l'exposé précédent. A chaque essai, on affichait une page de lettres minuscules choisies au hasard. Un signe "plus" (+) dans la marge de gauche indiquait la ligne sur laquelle on demandait au sujet de porter son attention. Nous avons choisi comme cibles de notre recherche les cinq lettres descendantes (g, j, p, q et y), parce que leur longueur constituait une variable intéressante. Une de ces cinq lettres se trouvait toujours au début de la ligne à balayer et indiquait au sujet la lettre cible au début de chaque essai. Le sujet devait alors balayer la ligne pour repérer cette lettre et, dès qu'il rencontrait, nommer immédiatement la lettre suivante. On demandait en outre au sujet de balayer la ligne une seule fois. S'il arrivait à la fin de la ligne sans

avoir pu identifier la lettre cible, on lui demandait de nommer la dernière lettre de la ligne. La lettre cible, de même que la lettre qui la suivait et la dernière lettre de la ligne, n'apparaissaient qu'une seule fois dans la ligne après l'indication donnée par la première lettre. On demandait aussi au sujet d'essayer de régulariser sa vitesse de balayage de façon à ne pas rater plus de 10 p. cent de ses essais. L'image apparaissait de façon relativement instantanée et la réponse donnée actionnait un interrupteur commandé par la voix qui effaçait l'image. On mesurait le temps de réponse à la milliseconde (ms) près et l'expérimentateur introduisait la réponse proprement dite au pupitre de commande pendant les trois secondes d'intervalle entre chaque essai.

Le sujet s'assoyait à environ quatre pieds de l'écran, la tête non retenue. L'expérimentateur affichait une page pour mieux expliquer l'expérience au sujet, puis il lui faisait faire 10 essais préliminaires. L'expérience en soi consistait en un plan de 27 mesures centrées sur le sujet, tenant compte de trois facteurs: la longueur des descendantes, l'espacement horizontal et l'espacement vertical. Par longueur des descendantes, on entend le degré de dépassement des caractères descendants sous la ligne, soit 0, 1 ou 2 éléments imagiers. L'espacement vertical et l'espacement horizontal étaient de 1, 2 ou 3 éléments imagiers. Le plan total comprenait donc 27 cellules. Chacune des cinq lettres cibles était présentée à deux reprises dans chaque condition d'expérimentation. Les 27 conditions apparaissaient au hasard dans chacun des 10 groupes d'essais. Le nombre total d'essais pour chacun des 11 sujets se chiffrait donc à 270. Étant donné que la session d'expérimentation durait environ 35 minutes, le sujet avait droit à une courte pause après le 135^{e} essai.

Nous avons effectué une analyse de variance des mesures répétées sur chacune des trois variables dépendantes suivantes: (1) la fréquence des erreurs, (2) la vitesse de balayage exprimée en caractères à la seconde, et (3) la vitesse de balayage exprimée en éléments imagiers à la seconde. L'analyse de la fréquence des erreurs a révélé un effet marqué de l'espacement horizontal ($F(2,20)= 3.595$, $p<.05$), et un certain effet de la longueur des descendantes ($F(2,20)= 29.378$, $p<.01$). Aucune interaction n'était significative. L'analyse de la vitesse de balayage exprimée en caractères à la seconde a révélé un effet de l'espacement vertical juste trop peu marqué pour servir de critère ($F(2,20)=3.311$, $p<.10$). Cependant, cette analyse a révélé un effet marqué de l'espacement horizontal ($F(2,20)= 30.208$, $p<.01$) et de la longueur des descendantes ($F(2,20)=16.514$, $p<.01$). Aucune interaction n'était significative. L'analyse de la vitesse de balayage exprimée en éléments imagiers à la seconde a révélé un effet marqué de l'espacement horizontal ($F(2,20)=77.701$, $p<.01$) et de la longueur des descendantes ($F(2,20)=16.361$, $p<.01$). Les figures 1 et 2 indiquent les moyennes correspondantes à ces effets. Le tableau 1 répartit la valeur de la variance totale pour chacun des effets marqués.

Tableau 1

Pourcentage de variance totale expliquée

	Variables		
	Fréquence des erreurs	Caractères à la seconde	Eléments imagiers à la seconde
Espacement horizontal	2,5	4,5	9,8
Longueur des descentantes	16,1	3,2	3,2

Nous avons vérifié la signification statistique des écarts entre les moyennes, à l'aide de la procédure Newman-Keuls (Winer, 1962). La fréquence des erreurs diminuait sensiblement avec chaque augmentation de la longueur des descendantes. De plus, la vitesse de balayage exprimée en caractères ou en éléments imagiers à la seconde augmentait sensiblement avec chaque augmentation de la longueur des descendantes. On peut déduire de ces constatations que la queue d'un caractère descendant du jeu de caractères 7 sur 9 devrait dépasser la ligne d'au moins deux éléments imagiers.

L'espacement horizontal entre les lettres a eu une incidence sur les deux mesures de la vitesse de balayage. Avec chaque augmentation de l'espacement, la vitesse de balayage exprimée en caractères à la seconde a diminué de façon notable, alors que la vitesse de balayage exprimée en éléments imagiers à la seconde augmentait. Ni l'une ni l'autre de ces mesures ne laisse présager l'approche d'une asymptote entre les deux extrêmes de la variable indépendante utilisée. De toute évidence, l'augmentation de la vitesse de balayage dont s'accompagnait l'augmentation de l'espacement des caractères ne suffisait pas à conserver à une valeur constante le taux d'introduction des caractères. Considérées isolément, ces données indiqueraient que l'espacement minimum est le meilleur, puisqu'il donnerait le taux d'introduction de l'information le plus élevé. Cependant, l'analyse de la fréquence des erreurs a montré que l'espacement de trois éléments imagiers occasionnait sensiblement moins d'erreurs que les deux autres espacements plus restreints, lesquels donnaient des résultats semblables. Il semble donc qu'un espacement horizontal d'un ou deux éléments imagiers donnent un taux d'introduction de caractères trop élevé. Nous préférons recommander le taux d'introduction plus lent que donne l'espacement de trois éléments imagiers, en raison de la fréquence d'erreurs plus faible qui en résulte. Ces données indiquent que l'on devrait espacer les caractères de trois éléments imagiers pour assurer l'identification juste des caractères.

L'espacement entre les lignes de lettres n'a eu d'influence marquée sur aucune des mesures dépendantes utilisées.

En résumé, notre expérience montre que les caractères descendants devraient dépasser la ligne d'au moins deux éléments imagiers, que l'espacement entre les caractères d'une ligne devraient être de trois éléments imagiers, et que l'espacement entre les lignes n'a pas d'influence sur l'identification des

lettres. Ces recommandations diffèrent des recommandations contenues dans le document précédent; elles montrent que la longueur des descendantes et l'espacement entre les caractères sur un écran cathodique dépendent du jeu de caractères que l'on emploie.

Ouvrage de référence

Winer, B.J. Statistical Principles in Experimental Design.

New York: McGraw-Hill, 1962.

FIGURE 1

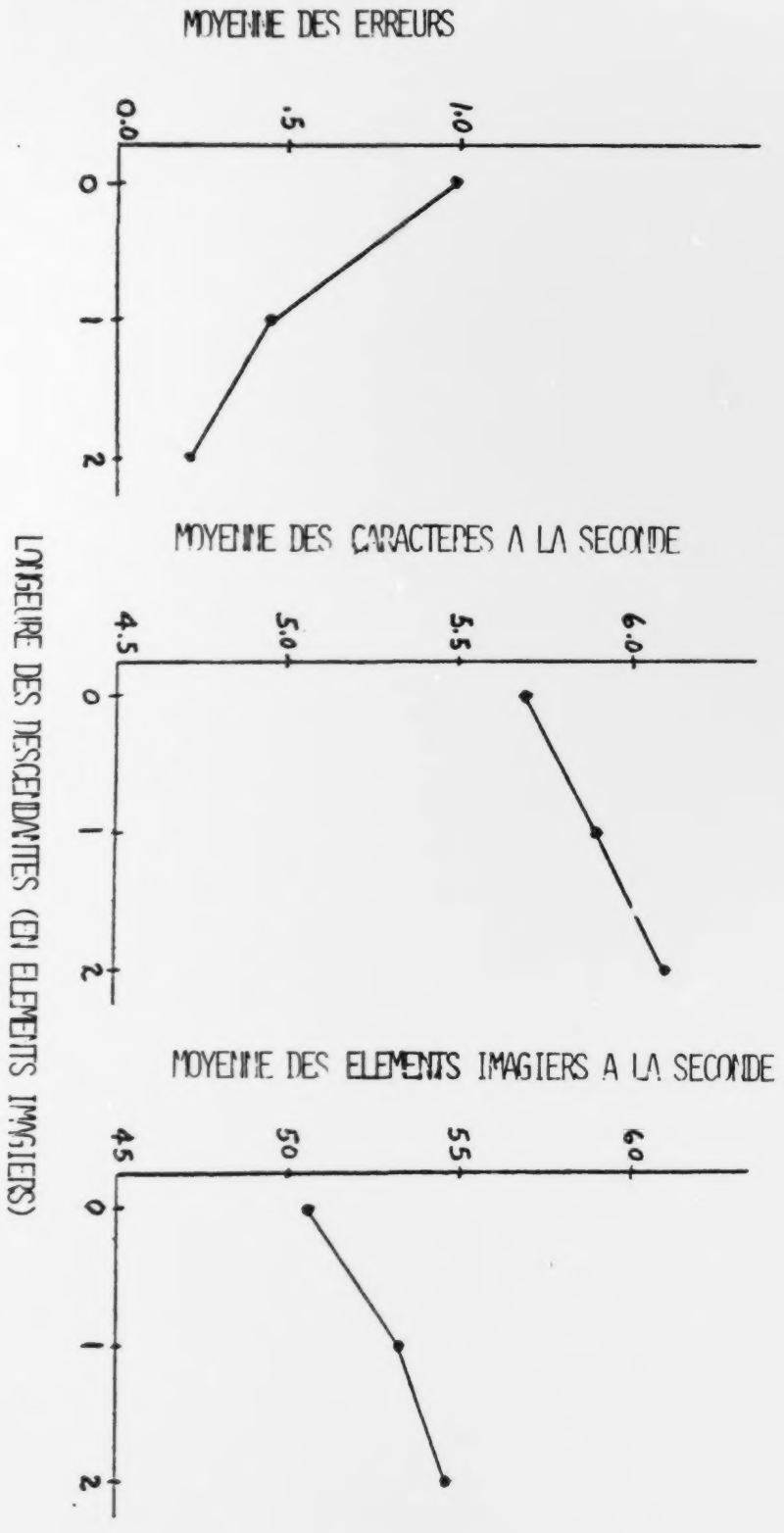
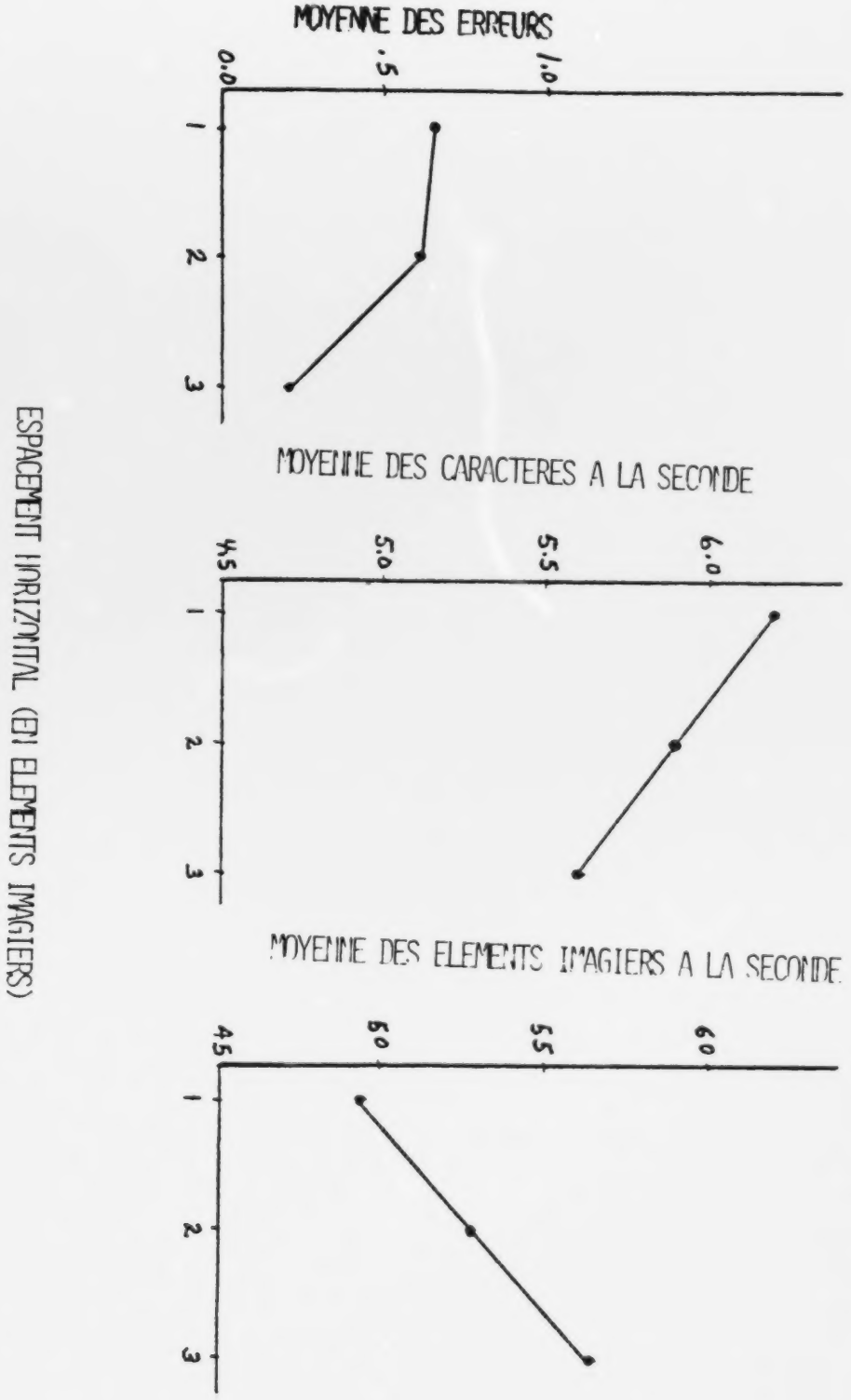

MOYENNE DES ERREURS
0.0
.5
1.0
0
1
2
MOYENNE DES CARACTERES A LA SECONDE
4.5
5.0
5.5
6.0
0
1
2
MOYENNE DES ELEMENTS IMAGIERS A LA SECONDE
4.5
5.0
5.5
6.0
0
1
2
LONGUEUR DES DESCENDANTES (EN ELEMENTS IMAGIERS)

FIGURE 2


MOYENNE DES ERREURS
0.0
.5
1.0
1
2
3
MOYENNE DES CARACTERES A LA SECONDE
4.5
5.0
5.5
6.0
1
2
3
MOYENNE DES ELEMENTS IMAGIERS A LA SECONDE
45
50
55
60
1
2
3
ESPACEMENT HORIZONTAL (EI ELEMENTS IMAGIERS)

Accentuation des caractères

W.C. Treurniet et S. Latrémouille

Ministère des Communications

Ottawa (Ontario)

Novembre 1979

Accentuation des caractères

William C. Treurniet et Susane Latrémouille
Communications Canada

Pour accentuer les lettres, on peut avoir recours à l'une ou l'autre des deux techniques suivantes: la méthode des ensembles et la méthode des composantes.

La méthode des ENSEMBLES. L'accent et la lettre à accentuer sont décrits sur une seule matrice de caractère et reçoivent un seul code. Cette méthode permet de comprimer verticalement la lettre afin de gagner l'espace nécessaire pour y insérer l'accent.

La méthode des COMPOSANTES. L'accent et la lettre à accentuer sont décrits sur des matrices de caractères distinctes. On accentue la lettre en superposant la matrice de l'accent sur la matrice de la lettre.

Les photographies des figures 1 à 4 montrent l'apparence des lettres accentuées selon la méthode des ENSEMBLES, et trois différentes applications de la méthode des COMPOSANTES.

Echantillons de la méthode des ENSEMBLES.

Figure 1. On utilise la méthode des ensembles pour comprimer les accents dans la matrice de caractère 5 sur 7. Lorsque cela est possible, on réduit la hauteur des majuscules comme des minuscules afin de laisser l'espace d'un élément imagier entre l'accent et la lettre à accentuer. On dispose tous les accents sur les deux lignes supérieures de la matrice.

Echantillons de la méthode des COMPOSANTES.

Figure 2. On utilise la méthode des composantes pour accentuer les lettres de la matrice de caractères 5 sur 7. Les lettres accentuées ont la même forme que les lettres sans accent. Avec cette méthode, il est impossible d'accentuer les lettres majuscules. On met tous les accents sur les deux lignes supérieures de la matrice.

Figure 3. On utilise la méthode des composantes pour mettre les accents dans l'espace contenu entre les lignes, large de trois éléments imagiers. Les accents ont une hauteur de deux ou de trois éléments imagiers. Les lettres accentuées ont la même grandeur et la même forme que les lettres sans accent.

Figure 4. On utilise la méthode des composantes pour mettre les accents dans l'espace contenu entre les lignes lorsqu'on accentue des lettres majuscules; ils empiètent sur la ligne supérieure de la matrice lorsqu'on accentue des caractères minuscules. La grandeur et la forme des lettres accentuées sont les mêmes que pour les lettres sans accent. Cette méthode oblige à distinguer entre les majuscules et les minuscules pour placer l'accent au bon endroit. Les accents ont une hauteur de deux ou de trois éléments imagiers.

Discussion:

La partie (a) de chaque figure montre les lettres accentuées les plus courantes de l'alphabet français, telles qu'elles apparaîtraient avec l'une ou l'autre des méthodes d'accentuation. Les parties (b) et (c) de chaque figure montrent les résultats de chaque méthode d'accentuation avec un texte en lettres minuscules et un texte en lettres majuscules. La figure 2(c) n'existe pas parce que les lettres majuscules ne sont pas accentuées dans ce cas.

Nous avons demandé à vingt francophones de mettre par ordre de préférence les parties (b) de chaque figure, du point de vue de l'accentuation des lettres (Annexe A). L'analyse de la variance de Friedman sur les choix de nos sujets a donné un chi au carré (X^2) de 37,8, qui est hautement significatif ($p<.001$). Ce résultat signifie que la cote donnée aux images différait. Nous avons comparé les cotes de chaque paire d'images, à l'aide de tests binomiaux. Nous avons établi le nombre de fois qu'une image était préférée à une autre; nous avons calculé ensuite la probabilité que cette fréquence de préférence s'écarte de la fréquence probable prévue. Chacune des six comparaisons par paires a montré une différence importante à plus du seuil de confiance de 95 p. cent.

L'analyse a montré que les techniques d'accentuation des figures 1 à 4 avaient reçu l'ordre de préférence décroissant suivant: figure 4, figure 3, figure 1 et figure 2. De plus, on ne peut considérer deux méthodes comme équivalentes, en termes de préférence.

Les lettres accentuées des figures 2, 3 et 4 sont conçues pour pouvoir être accentuées à l'aide de la méthode des composantes. On doit cependant noter qu'on peut également accentuer ces lettres à l'aide de la méthode des ensembles. A cette fin, on n'a qu'à étendre verticalement la grandeur de la matrice de caractère des figures 3 et 4; en somme, on doit faire passer la matrice de caractère 5 sur 7 à une matrice de caractère 5 sur 10 de façon à y inclure l'espace de trois éléments imagiers qui sépare les lignes.

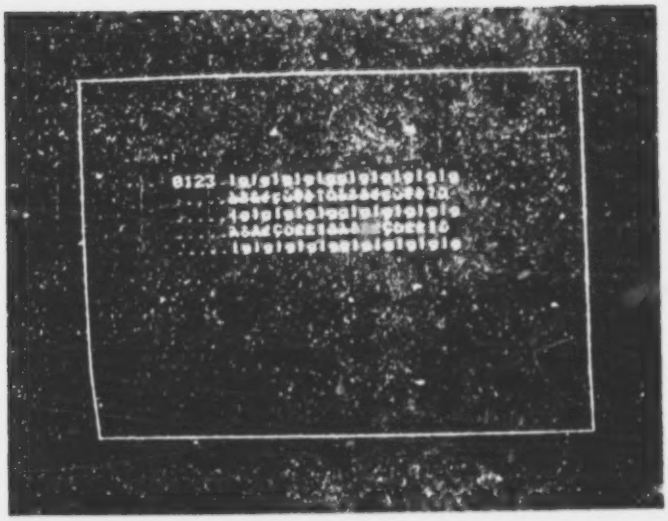


Figure 1(a)

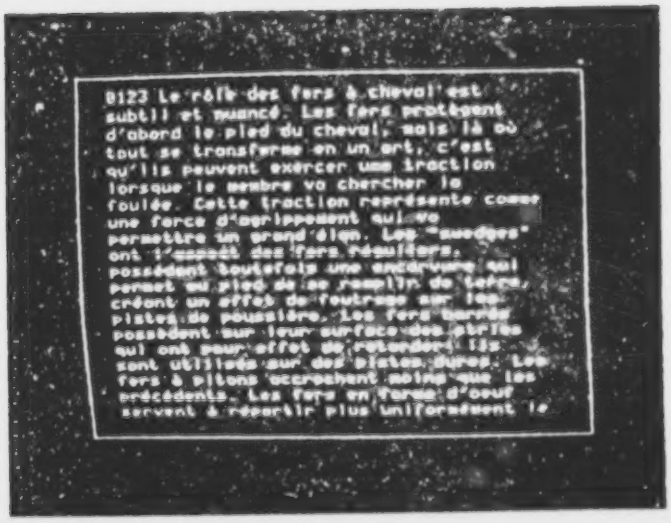


Figure 1 (b)

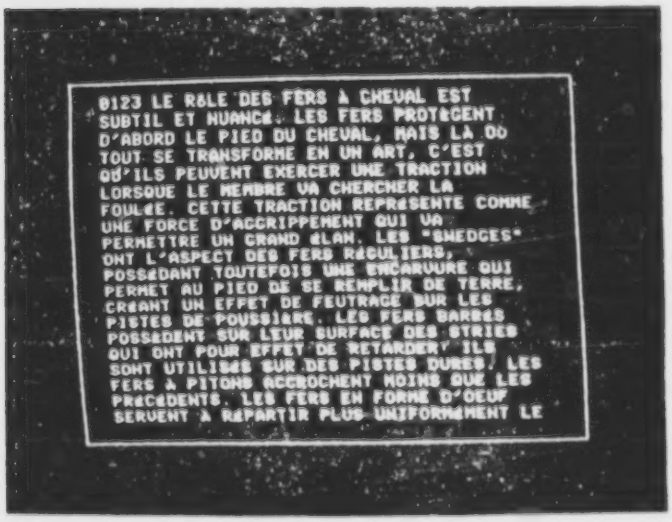


Figure 1 (c)

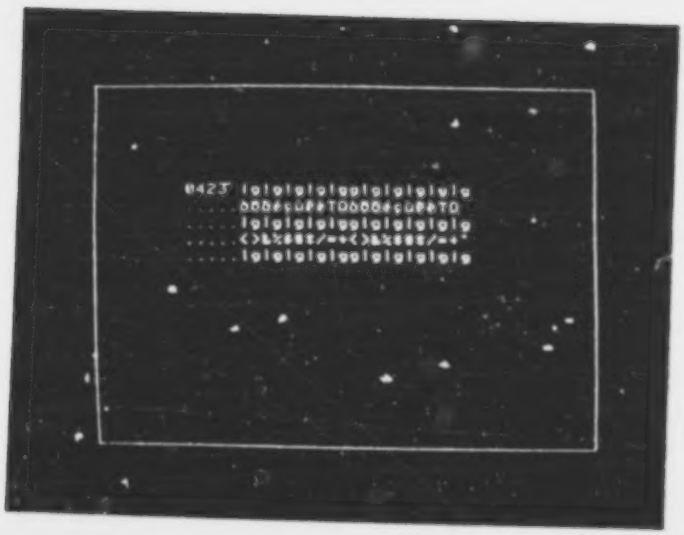


Figure 2 (a)

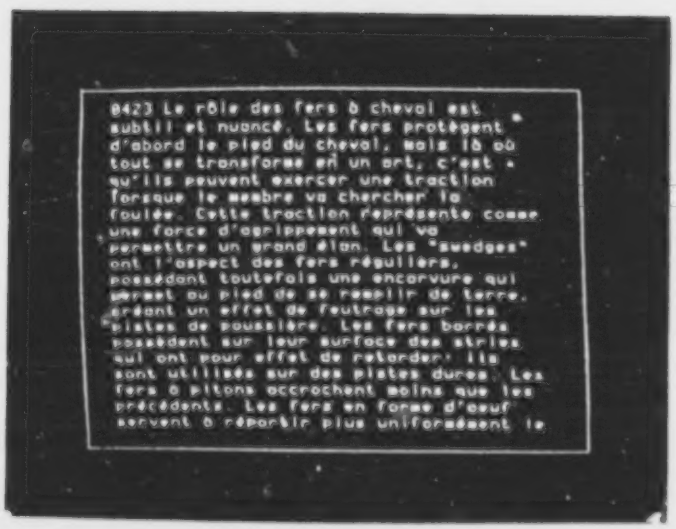


Figure 2 (b)

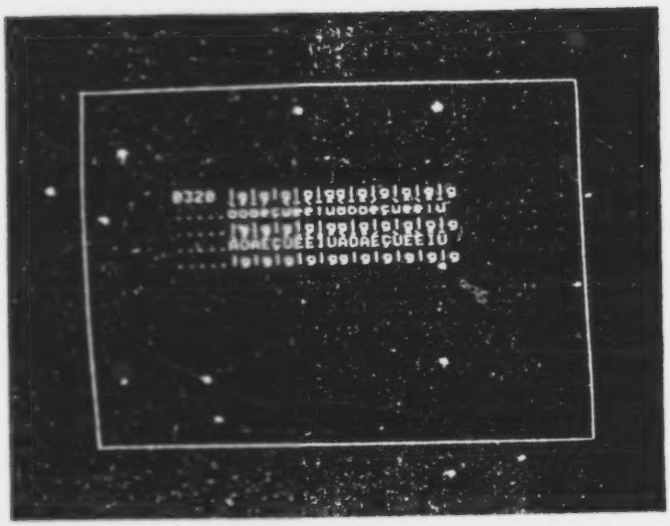


Figure 3 (a)

Figure 3 (b)

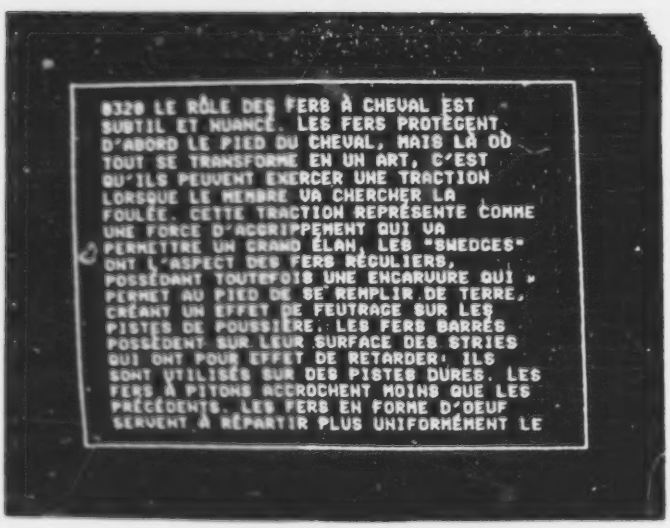
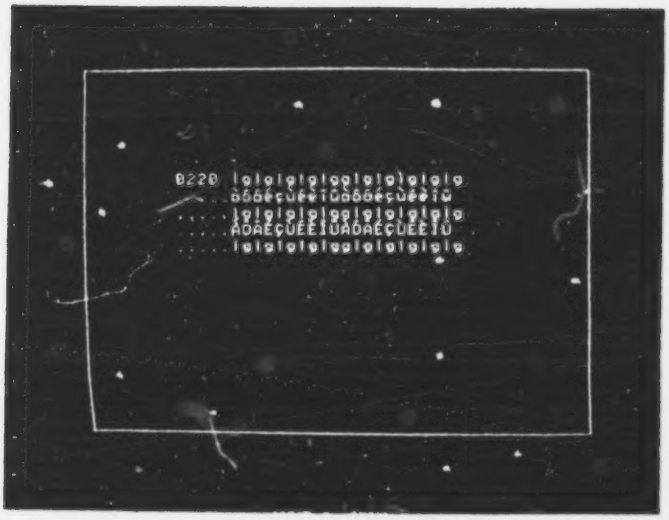


Figure 3 (c)


0220 Ie!e!e!e

Annexe A

Chaque sujet devait attribuer la cote "1" à l'image qu'il aimait le plus et la cote "4" à celle qu'il aimait le moins. Le tableau ci-dessous indique combien de fois chaque image a reçu chacune des quatre cotes possibles:

Cote	Figure 1	Figure 2	Figure 3	Figure 4
1	1	1	1	17
2	4	0	13	3
3	10	6	4	0
4	5	13	2	0

CHAPITRE III

FACTEURS HUMAINS RATTACHES AUX BASES DE DONNEES.

NOMBRE OPTIMAL DE RUBRIQUES QUE DEVRAIT PROPOSER UNE PAGE D'INDEX DANS LA BASE DE DONNEES DU SYSTEME TELIDON INTERACTIF

ERIC LEE
GROUPE DE RECHERCHE EN COMPORTEMENT
MINISTERE DES COMMUNICATIONS
OTTAWA

Décembre 1979

RESUME

La détermination du nombre optimal de rubriques que devrait proposer une page d'index du Télidon donne lieu à plusieurs constatations secondaires d'ordre pratique. Le présent document vise les objectifs suivants:

a) analyser la question du point de vue du comportement de l'utilisateur;

b) élaborer un modèle destiné à prédire les effets de la variation du nombre de rubriques proposées par page d'index; et

c) décrire les conclusions pratiques qu'on peut tirer de la présente analyse.

A l'intention des lecteurs qui ne seraient pas intéressés à connaître tous les détails de l'analyse, nous en esquissons ici le cheminement logique, suivi des principales conclusions qui en découlent. L'analyse détaillée et la discussion des hypothèses et des calculs mathématiques se retrouvent dans la section intitulée "ANALYSE".

nécessaire pour repérer et enfoncer les touches appropriées sur le clavier et c) du temps qu'il faut à l'ordinateur pour répondre à la demande et afficher la page d'index suivante. C'est l'importance relative de ces trois facteurs qui déterminera si le temps de recherche total est plus long ou moins long lorsqu'on augmente le nombre de rubriques par page.

Le temps requis pour trouver l'information dans la base de données du Télidon, c'est-à-dire le temps de recherche, sera probablement un élément déterminant de l'attrait du système et de la satisfaction de l'utilisateur. Il est permis de présumer qu'on refusera d'utiliser un système de recherche documentaire les temps de recherche ST sont d'une longueur excessive. Ainsi, plus on écourtera le temps de recherche moyen, plus le système sera bien accueilli et donnera satisfaction aux utilisateurs et plus les gens seront enclins à recourir aux services du Télidon.

Les temps de recherche totaux qu'exigera le système dépendront de nombreux facteurs comme le nombre de documents dans la base de données, le nombre de niveaux de l'index ramifié, le nombre de rubriques proposées par chaque page d'index, le temps de réponse du système Télidon, la vitesse de lecture des rubriques, le temps nécessaire pour repérer et enfoncer les touches et la méthode de recherche de chaque utilisateur. Des prédictions assez précises peuvent être faites quant aux temps de recherche à partir d'hypothèses raisonnables concernant à la fois la vitesse de lecture des utilisateurs et les méthodes qu'ils emploient

La version du Télidon qui servira aux expériences-pilotes et au lancement initial du système sur le marché prévoit la consultation d'un index ramifié comme principale méthode de recherche documentaire que pourront adopter les utilisateurs. Pour repérer un document dans la base de données au moyen de cette méthode, l'utilisateur devra faire un choix parmi les rubriques apparaissant sur chacune des pages d'index qui lui seront présentées successivement. Dans le cas d'une base de données dont le volume est fixe (c'est-à-dire une base de données contenant un nombre déterminé de documents), le nombre de pages d'index que doit consulter l'utilisateur pour repérer les documents qu'il cherche dépend du nombre de rubriques proposées par chaque page d'index, et vice-versa. Lorsqu'on augmente le nombre de rubriques par page d'index, on réduit le nombre de pages d'index à consulter pour retracer un document.

Dès lors qu'il y a moins de pages d'index à consulter pour retracer un document, le temps total consacré à l'obtention des pages d'index est également réduit. Du même coup, par contre, en augmentant ce nombre de rubriques par page, on augmente le nombre de rubriques à lire à chaque page et, par conséquent, le temps que mettront les utilisateurs à consulter chaque page d'index. L'effet global du nombre de rubriques par page sur le temps total, ci-après appelé ST ou temps de recherche, que mettra un utilisateur à repérer l'information voulue (c'est-à-dire un document) dans la base de données du Télidon, dépendra a) du temps qu'il faut pour lire les rubriques de chaque page d'index, b) du temps

pour consulter les rubriques affichées sur une page d'index, et des projections de Bob Baser concernant les temps de réponse du système Télidon (voir note ci-jointe). L'analyse présentée dans la section suivante décrit un modèle de comportement élaboré à cette fin et montre comment il permet de résoudre le problème très concret du nombre optimal de rubriques que devrait proposer chaque page d'index. On notera que si le nombre optimal de rubriques excédait 10, on aurait là un argument de taille pouvant justifier l'adoption d'un code alphabétique (permettant d'employer jusqu'à 26 lettres) pour identifier les rubriques. Dans ce cas, il faudrait, entre autres choses, fournir à chaque utilisateur un clavier non plus strictement numérique, mais alphanumérique, ce qui serait plus coûteux.

Les principales conclusions de l'analyse (décrites plus en détail dans la section suivante) sont les suivantes:

1) Le nombre optimal de rubriques par page d'index du Télidon (propre à réduire le plus possible le temps que les utilisateurs doivent consacrer à la recherche par arborescences) est inférieur à 10 dans un très grand nombre de conditions possibles (voir les tableaux 1 et 2).
2) Il devrait toujours y avoir, si possible, de 4 à 8 rubriques par page, c'est-à-dire le nombre optimal de rubriques dans la plupart des cas (voir les tableaux 1 et 2).
3) Les fournisseurs d'information devraient éviter, si possible, d'afficher seulement deux rubriques par page d'index parce que les temps de recherche correspondant à deux rubriques par page sont beaucoup plus longs que le temps optimal (voir les figures 1 et 2).

Nous recommandons:

QU'ON ADOPTE POUR LE TELIDON UN INDEX RAMIFIE A PAGES NUMERIQUEMENT IDENTIFIEES DONT CHACUNE COMPRENDRAIT UN MAXIMUM DE DIX RUBRIQUES, DE PREFERENCE A UN INDEX RAMIFIE A PAGES ALPHABETIQUEMENT IDENTIFIÉES DONT CHACUNE COMPRENDRAIT UN MAXIMUM DE 26 RUBRIQUES.

ANALYSE

Les observations préliminaires du comportement des utilisateurs du Télidon montrent que ceux-ci peuvent recourir à deux méthodes cognitives distinctes pour chercher une rubrique dans une page d'index: la recherche exhaustive (RE)et la recherche autocessante (RA). Dans le premier cas (RE), l'utilisateur lit toutes les rubriques d'une page d'index avant de faire le choix qui déclenche l'apparition de la page d'index suivante. Dans le second cas (RA), par contraste, sa lecture des rubriques cesse dès qu'il en trouve une qui lui apporte l'élément de réponse voulu pour passer à la page d'index suivante. L'utilisateur décide d'accepter ou de rejeter chacune des rubriques qu'il rencontre. D'après ce que j'ai eu l'occasion d'observer du comportement des utilisateurs, il semble qu'environ 75% ont recours à la méthode de recherche autocessante. (Naturellement, une recherche plus systématique serait nécessaire pour déterminer exactement la proportion d'utilisateurs qui ont recours à chaque méthode). Quoi qu'il en soit, nous allons voir que le nombre optimal de rubriques par index est presque toujours inférieur à 10, quelle que soit la méthode utilisée.

Lorsqu'ils consultent une page d'index, tous les utilisateurs semblent parcourir les rubriques de façon ordonnée, du haut vers le bas de la page. Celui qui a recours à la méthode de recherche exhaustive lit toujours la totalité des a rubriques que contient une page d'index. Le nombre prévu de rubriques par page que

cet utilisateur doit consulter est donc simplement E(A) = a. Au début, tout au moins, les rubriques des pages d'index y seront probablement placées au hasard. Par conséquent, en supposant que les a rubriques d'une page y soient réparties au hasard, la méthode de recherche autocessante exigerait, à la longue, que l'utilisateur lise en moyenne la moitié des rubriques de chaque page d'index avant de rencontrer celle qu'il cherche. Autrement dit, E(A) = a/2.

De même, lorsque les rubriques sont classées par ordre alphabétique (pour une sélection à peu près au hasard des rubriques voulues et une répartition à peu près égale des choix exacts de rubriques), E(A) = a/2 pour la méthode de recherche autocessante. (Si les rubriques les plus souvent consultées se retrouvent à peu près en début d'index, alors E(A)→1. D'autre part, si elles se retrouvent à peu près en fin d'index, E(A)→a. Cependant, dans une base de données contenant un nombre important de pages d'index, il est très improbable que les rubriques les plus souvent consultées commencent toutes par z ou par a.)

Evidemment, l'expérience aidant, les utilisateurs pourraient se souvenir de la position occupée par la rubrique recherchée et n'auraient alors qu'à regarder les premières lettres des rubriques disposées par ordre alphabétique pour retrouver le numéro d'index correspondant.

Il est certainement beaucoup plus rapide de jeter un coup d'oeil, sur la première lettre des rubriques que de lire les rubriques en entier. Ainsi, il pourrait suffire à un habitué de lire deux ou trois

rubriques dans une page pour y retrouver celle qu'il recherche. Dans ce cas, E(A) ne serait plus égal à a/2 pour une recherche autocessante, mais tendrait vers 1. Cependant, il sera extrêmement difficile de se souvenir de tous les termes d'index du Télidon en raison de la dimension même des bases de données projetées. En outre, de nombreuses expériences psychologiques (voir n'importe quel texte d'introduction sur le sujet) ont amplement démontré qu'il est beaucoup plus facile de reconnaître des mots déjà vus que de se les rappeler. Finalement, il n'est sûrement pas risqué de prédire que la majorité des utilisateurs privés à qui s'adresse le Télidon ne deviendront jamais des experts dans la consultation des index. Si le classement alphabétique n'a aucun effet sur le nombre prévu de rubriques par page que l'utilisateur devra lire, l'effet d'un classement par fréquence de consultation pourrait être spectaculaire. Les rubriques d'une page d'index peuvent y être affichées dans l'ordre correspondant à la fréquence avec laquelle les utilisateurs les ont consultées antérieurement, la rubrique la plus souvent consultée figurant en première place sur la page et la moins souvent consultée, au bas de la page. Dans le cas extrême ou les utilisateurs choisiraient presque toujours une rubrique donnée, celle-ci serait placée en tête de liste et donnerait une recherche autocessante donnerait $E(A) \to 1$. Si, au contraire, toutes les rubriques sont choisies un nombre égal de fois, leur agencement ou classement se ferait essentiellement au hasard et on obtiendrait $E(A) = a/2$ pour une recherche autocessante. Donc, en pratique, la réduction du temps de recherche total résultant

du classement par fréquence des rubriques d'une liste profite seulement aux utilisateurs qui consultent selon la méthode de recherche autocessante des pages d'index dont les rubriques n'ont pas été consultées un nombre égal de fois dans le passé.

La relation entre le nombre de documents d'une base de données (représenté par $\underline{d}$), le nombre de rubriques par niveau ou page de l'index ramifié (représenté par $\underline{a}$) et le nombre de niveaux de l'index de la base de données (représenté par $\underline{p}$ qui correspond également au nombre de pages d'index que l'utilisateur doit consulter pour repérer chaque document) est donnée par l'équation suivante:

1) $d = a^p$.

Si l'on connaît la valeur de deux des trois variables $\underline{a}$, $\underline{d}$, $\underline{p}$, la troisième est fixe et peut être calculée par l'équation 1. Par exemple, pour une base de données de 100 000 documents et 10 rubriques par niveau, le nombre de niveaux est

2) $p = \log d / \log a$

$= 5/1 = 5.$

Pour retracer un document dans la base de données, l'utilisateur doit consulter au total $\underline{p}$ index (une page par niveau de la structure). Le temps nécessaire pour avoir accès à chaque page et pour en lire les rubriques (le nombre moyen de rubriques E(A) que l'utilisateur doit lire par page varie selon la méthode de recherche) est:

3) $T = E(A) \cdot t_a + t_{kp} + t_{rt}$

où E(A) = nombre prévu de rubriques à consulter par page pour trouver la rubrique voulue;

t_a = temps requis pour lire une rubrique (temps de lecture de l'utilisateur);

t_{kp} = intervalle entre la lecture des rubriques d'une page d'index et le moment où l'utilisateur enfonce les touches;

t_{rt} = temps de réponse de l'ordinateur, c'est-à-dire le temps qui s'écoule entre le moment où la dernière touche est enfoncée et le moment où la page d'index suivante est affichée.

Ainsi, le temps de recherche total est simplement égal à p fois l'équation 3 ou

$$4) \quad ST = P(E(A) \cdot t_a + t_{kp} + t_{rt})$$

où ST = temps de recherche total;

p = nombre de niveaux et de l'index ramifié.

Par substitution de $d = a^p$, on obtient

$$5) \quad ST = \frac{E(A) \cdot t_a + t_{kp} + t_{rt}}{\log a} \log d$$

où d = nombre de documents dans la base de données et

a = nombre de rubriques par page d'index.

L'équation 5 montre que le nombre de documents de la base de données (représenté par log d) n'influe pas sur le nombre optimal de rubriques. Cette constatation est illustrée par la figure 1 qui représente graphiquement la relation entre le nombre de rubriques par page, a, et le temps de recherche total (ST). Les temps de recherche ST portés sur le graphique de la figure 1 correspondent à une base de données comprenant 10 documents (donc, log d = 1).

Pour obtenir les temps de recherche ST correspondant à une base de données constituée de 100 documents (log d = 2), il faut multiplier tous les temps ST de la figure 1 par la constante log d = 2. Pour trouver graphiquement la relation entre ST et a d'une base de données d = 100, il suffit de doubler toutes les valeurs en ordonnée (c'est-à-dire les valeurs de ST). On remarquera que le tracé résultant serait identique à celui de la figure 1, sauf que les valeurs en ordonnée seraient doublées. Le nombre optimal de rubriques par page ne suit pas les changements de d. Autrement dit, le point le plus bas de chaque courge de la figure 1 demeure au niveau le plus bas lorsque la valeur de d varie. Par conséquent, dans les calculs ultérieurs, la dimension de la base de données est fixée arbitrairement à 10 et $\log_{10} d = 1$. L'expression log d disparaît alors de l'équation 5. (En fait, la dimension de la base de données influe sur la valeur de ST. Le temps de recherche correspondant à une base de dimension donnée s'obtient facilement en multipliant par log d les temps de recherche signalés.

Ainsi, le temps de recherche d'un document d'une base de données contenant exactement 10 documents s'obtient directement à partir de la figure 1. Dans le cas d'une base de données qui contient 100 000 documents, le temps de recherche (ST) d'un document s'obtient en multipliant par log 100 000 = 5 la valeur appropriée de ST de la figure 1.)

RESULTATS

Pour toute combinaison donnée de nombre d'utilisateurs, méthode de recherche et temps de lecture, il existe un nombre optimal de rubriques que devrait afficher chaque page d'index pour réduire le temps de recherche total au minimum (voir les figures 1 et 2). Les tableaux 1 et 2 présentent le nombre optimal de rubriques correspondant à un large éventail de situations possibles.

Par exemple, des observations pratiques indiquent qu'il faut généralement de 0,5 à 1 s pour repérer et enfoncer les touches appropriées d'un clavier. La figure 1 et le tableau 1 présentent respectivement des temps de recherche prévus et des nombres optimaux de rubriques escomptés à partir de l'hypothèse qu'il faut, en moyenne, 0,5 s pour repérer et enfoncer les touches. Les prédictions présentées à la figure 2 et au tableau 2 se fondent sur l'hypothèse qu'il faut, en moyenne, 1 seconde.

On a fait varier systématiquement de 1 à 120 le nombre de personnes qui utilisent le Télidon en même temps. Les temps de réponse correspondants du Télidon, qui dépendent du nombre d'utilisateurs, varient du temps optimal prévu par Bob Baser à un temps de réponse très lent. Les temps de lecture par rubrique présentés aux tableaux 1 et 2 varient d'un temps théorique minimal de 0,25 s par rubrique (un mouvement saccadé de l'oeil durant en moyenne de 200 à 250 ms est ici nécessaire pour passer d'une rubrique à l'autre) à un temps très lent de 2 s par rubrique (temps qui équivaut à peu près à la vitesse de lecture très lente de 30 à 60 mots à la minute). Bill Treurniet estime que la vitesse de lecture moyenne se situe à environ 240 mots à la minute, ce qui correspond à peu près à un temps de lecture de 0,5 s par rubrique. Finalement, les tableaux tiennent compte à la fois de la méthode de recherche exhaustive et de la méthode de recherche autocessante.

L'étude des tableaux 1 et 2 montre que le nombre optimal de rubriques est inférieur à 10 dans un très grand nombre de conditions possibles. Il est supérieur à dix rubriques par page uniquement lorsque les lecteurs sont très rapides et que le système est très surchargé (c'est-à-dire lorsque plus de 100 utilisateurs effectuent simultanément une recherche documentaire, ce qui fait augmenter passablement le temps de réponse de l'ordinateur).

Même dans ces conditions, les temps de recherche obtenus se rapprochent du temps optimal si chaque page contient dix rubriques. Sur le plan pratique, cela signifie que l'on devrait adopter pour les bases de données du Télidon une structure ramifiée à pages d'index numériques d'un maximum de dix rubriques chacune, de préférence à des pages d'index alphabétiques d'un maximum de 26 rubriques chacune. Il devrait y avoir, si possible, à peu près 48 rubriques par page, soit le nombre optimal dans la plupart des situations (voir les tableaux et les figures 1 et 2).

Même si les temps de recherche correspondant à 10 rubriques par page ou moins sont plus courts (dans la plupart des cas) que ceux qui correspondent à 26 rubriques par page, l'inverse peut se révéler exact dans certains cas. Lorsque les rubriques sont affichées sur une page d'index dans l'ordre correspondant à leur fréquence de sélection et que la probabilité ou fréquence relative de sélection s'approche de l'unité, le temps de recherche devrait, en fait, être plus court dans le cas de 26 rubriques par page. Cette relation s'exprime par la formule suivante:

$$6) \quad ST_{a=2} = \left(\frac{\log 10}{\log 26}\right) ST_{a=10}$$

$$= 0,7067\ ST_{a=10}$$

Dans ces conditions, une économie de 30% du temps de recherche total est réalisable si l'on opte pour 26 rubriques par page au lieu de 10. En fait, dans ces conditions, la durée de la recherche est inversement proportionnelle au nombre de rubriques par page d'index. Dans la pratique, la limite semble être de 40 rubriques par page (deux rangées de 20 rubriques chacune). Ainsi, l'économie la plus importante du temps de recherche total serait:

$$7) \quad ST_{a=40} = \left(\frac{\log 10}{\log 40}\right) ST_{a=10}$$

$$= 0,6242\ ST_{a=10}$$

Cependant, il ne faut pas oublier que ces conditions ne se retrouveront vraisemblablement pas dans la pratique. Jusqu'à maintenant, notre analyse supposait l'affichage simultané de tous les renseignements textuels d'une page. Trois formes d'affichage peuvent en fait convenir aux systèmes Télidon: a) affichage instantané de chaque page, b) affichage du texte ligne par ligne à une vitesse supérieure à la vitesse de lecture de l'utilisateur et c) affichage du texte ligne par ligne à une vitesse inférieure à la vitesse de lecture de l'utilisateur. Les étapes de l'analyse exposées jusqu'à maintenant s'appliquent aux affichages a) et b).

Lorsque le texte est affiché ligne par ligne à une vitesse supérieure à la vitesse de lecture de l'utilisateur, on peut considérer que le texte apparaît en une seule fois parce que l'affichage ne réduit pas la vitesse de lecture (qui est l'un des paramètres de l'analyse présentée jusqu'ici). Cependant, lorsque le texte est affiché ligne par ligne à une vitesse inférieure à la vitesse de lecture de l'utilisateur, il y a forcément ralentissement de celle-ci. Dans ce cas, la vitesse de lecture des rubriques t_a doit être égale à la vitesse d'affichage t_d. Autrement dit, la vitesse de lecture maximale ne peut excéder la limite imposée par la vitesse d'affichage t_d. Par conséquent, on peut aussi examiner le cas c) au moyen de la présente analyse en posant t_a (vitesse de lecture) = t_d (vitesse d'affichage).

La vitesse d'affichage actuelle du Télidon est de 1200 bauds ou 120 caractères/s (voir note de Bob Baser ci-jointe). D'après un échantillon de 25 rubriques prélevé dans la base de données actuelle, le nombre de caractères par rubrique d'index serait de 13,66 (ce qui équivaut à peu près à deux mots par rubrique). Ce nombre comprend deux caractères pour le numéro de l'index associé à chaque rubrique et un caractère pour l'instruction de description de l'image (IDI) à la fin de chaque rubrique. Ainsi, la vitesse d'affichage t_d égale 1/10s/rubrique (120/13,66 = 8,78 rubriques/s), ce qui dépasse la vitesse maximale théorique de lecture de 1/4 s/rubrique. On peut donc considérer que chaque page est affichée instantanément et appliquer la présente analyse au système Télidon actuel.

Nous supposons ici, que le temps de réponse de l'ordinateur demeurera sensiblement le même malgré l'augmentation du volume de la base de données. Cette suppostion se révèlera à peu près exacte si les disques utilisés pour stocker les données sont presque remplis. Les temps de recherche de l'ordinateur demeurent relativement les mêmes pour des disques de différentes capacités. Ainsi, même si l'ordinateur doit consulter six fois plus de données sur un disque de 180 millions d'octets que sur à un disque de 30 millions d'octets, sont temps moyen de réponse sera à peu près le même. Cependant, si un disque de très grande capacité est utilisé pour stocker un nombre restreint de données (30 millions d'octets sur un disque de 180 millions d'octets, par exemple), le temps de réponse de l'ordinateur sera quelque peu plus rapide que le temps prévu sur le graphique de Bob Baser. Les résultats et conclusions de la présente analyse ne s'appliquent pas lorsque le volume de la base de données est très inférieur à la capacité de mémoire du disque. Néanmoins, dans ce cas, votre équation 5 peut servir à faire des prédictions en substituant aux temps de réponse de l'ordinateur les valeurs appropriées.

Comme dans tout modèle physique ou économique, la justesse des prédictions dépend de la justesse des hypothèses. Des recherches empiriques peuvent et doivent être faites pour confirmer les hypothèses du présent modèle.

TABLEAU 1

NOMBRE OPTIMAL DE RUBRIQUES PAR INDEX DU TELIDON

LORSQU'IL FAUT UNE DEMI-SECONDE POUR ENFONCER LES TOUCHES (t_{kp}=1/2)[a]

METHODE DE RECHERCHE	TEMPS DE LECTURE PAR RUBRIQUE	NOMBRE D'UTILISATEURS (TEMPS DE REPONSE DU TÉLIDON[b])			
		1-60(.55)	80(.60)	100(.90)	120(1.35)
EXHAUSTIVE	.25 sec	6	6	6/7[c]	7
	.50	4	4	5	5/6
	1.00	4	4	4	4
	2.00	3	3	3	4
AUTOCESSANTE	.25 sec	8	8	9	11
	.50	6	6	6/7	7
	1.00	4	4	5	5/6
	2.00	4	4	4	4

[a]Le nombre optimal de rubriques par page ne varie pas selon les variations du volume de la base de données (c'est-à-dire le nombre total de documents). Autrement dit, les valeurs optimales du tableau ci-dessus s'appliquent à n'importe quel nombre de documents.

[b]Mesurés en secondes. Ces temps de réponse du Télidon, t_{rt}, sont établis d'après les prédictions de Bob Baser (juillet 1979) pour le système Télidon qui sera bientôt commercialisé.

[c]Les valeurs 6 et 7 représentent toutes deux des valeurs optimales.

TABLEAU 2

NOMBRE OPTIMAL DE RUBRIQUES PAR INDEX DU TELIDON

LORSQU'IL FAUT UNE SECONDE POUR ENFONCER LES TOUCHES (t_{kp}=1)[a]

METHODE DE RECHERCHE	TEMPS DE LECTURE PAR RUBRIQUE	NOMBRE D'UTILISATEURS (TEMPS DE REPONSE DU TELIDON[b])			
		1-60(.55)	80(.60)	100(.90)	120(1.35)
EXHAUSTIVE	.25 sec	7	7	7/8	8
	.50	5	5	5	6
	1.00	4	4	4	5
	2.00	3	3	4	4
AUTOCESSANTE	.25 sec	10	10	11	12/13
	.50	7	7	7/8	8
	1.00	5	5	5	6
	2.00	4	4	4	5

[a]Le nombre optimal de rubriques par page ne varie pas selon les variations du volume de la base de données (c'est-à-dire le nombre total de documents). Autrement dit, les valeurs optimales du tableau ci-dessus s'appliquent à n'importe quel nombre de documents.

[b]Mesurés en secondes. Ces temps de réponse du Télidon, t_{rt}, sont établis d'après les prédictions de Bob Baser (juillet 1979) pour le système Télidon qui sera bientôt commercialisé.

Figure 1. Temps de recherche en fonction du nombre de rebriques par page d'index pour une méthode de recherche exhaustive

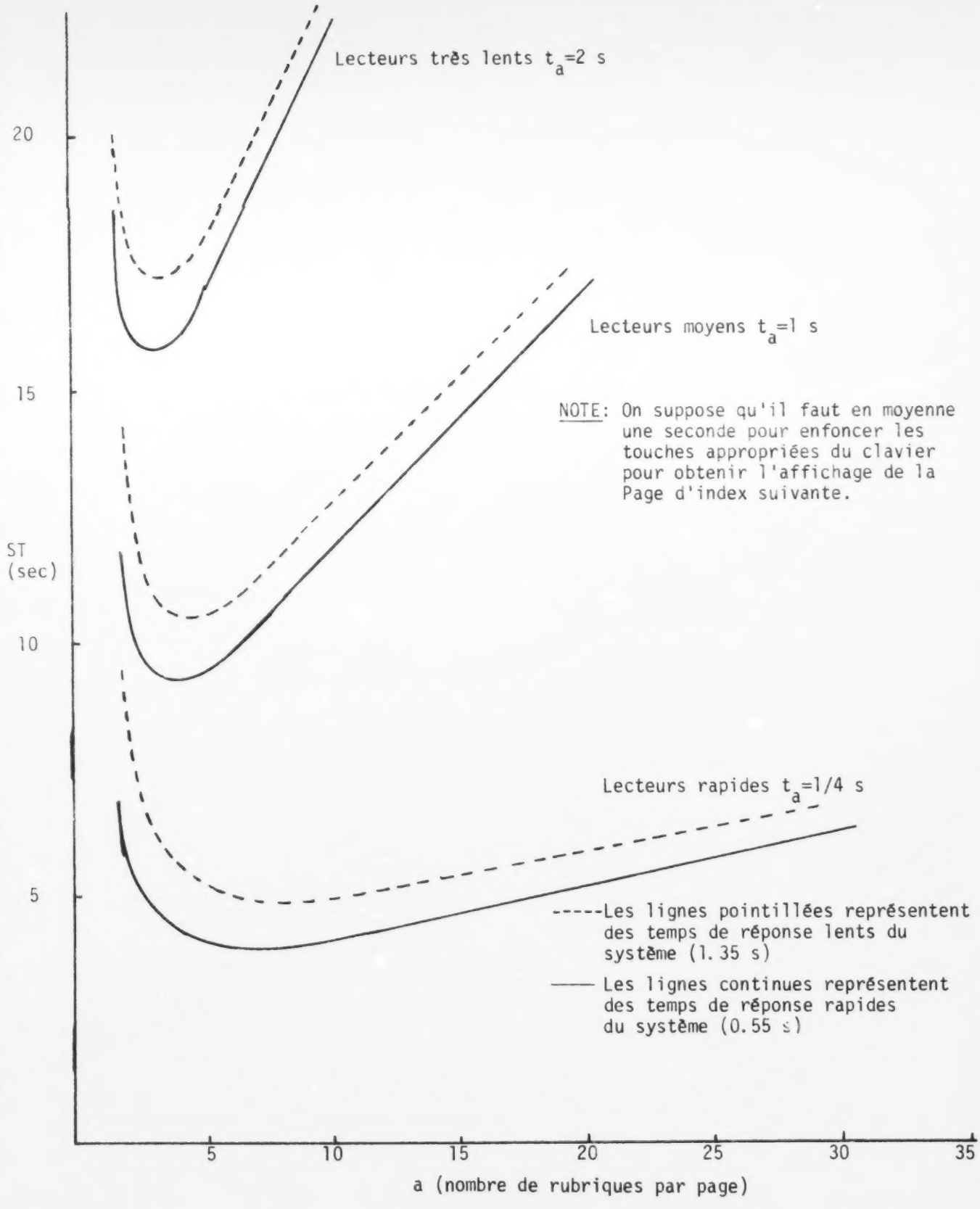
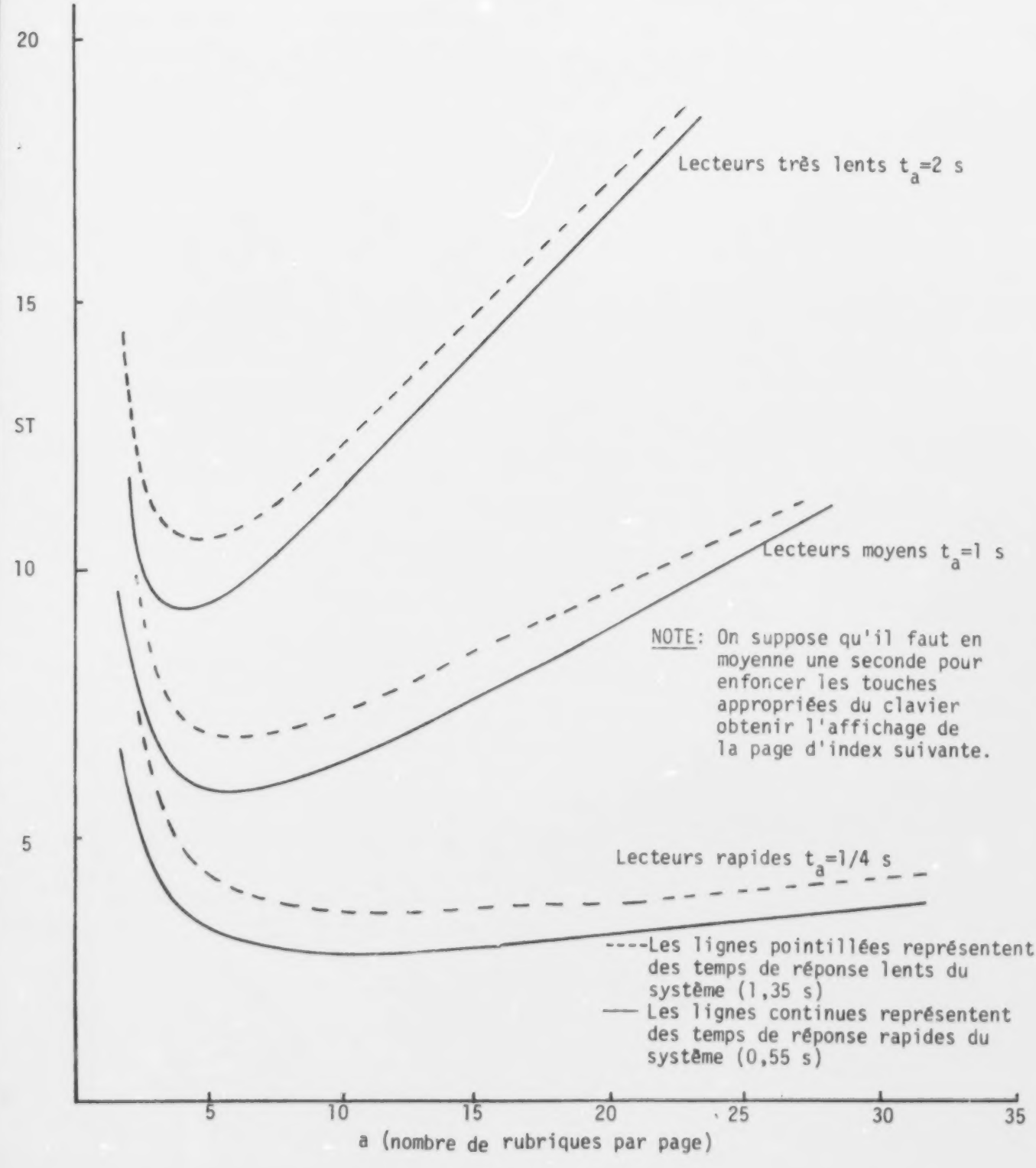


Figure 2. Temps de recherche en fonction du nombre de rubriques par page d'index pour une méthode de recherche exhaustive



Government of Canada / Gouvernement du Canada

MEMORANDUM **NOTE DE SERVICE**

TO / A: Eric Lee

FROM / DE: Robert Baser

SECURITY CLASSIFICATION / DE SÉCURITÉ:

OUR FILE / NOTRE RÉFÉRENCE:

YOUR FILE / VOTRE RÉFÉRENCE:

DATE: 5 juin 1979

SUBJECT / OBJET: Temps de réponse de la base de données du Télidon

Vous trouverez ci-joint le graphique des temps de réponse de notre base de données actuelle établi à partir de calculs théoriques.

Le temps de réponse du système est le temps qui s'écoule entre le moment où l'utilisateur du terminal enfonce la touche o et celui où le premier caractère d'une page est affiché sur l'écran. Le temps d'affichage n'a pas été pris en considération, mais on peut l'obtenir de la façon suivante:

256 caractères

(2,13 s) 0,25 s ....

Des blocs de 256 caractères sont transmis en 2,13 secondes avec un intervalle de 0,25 seconde entre chaque bloc.

Une analyse initiale montre que l'on pourrait prévoir les intervalles comme suit:

- bouclage de ligne (0,25 s à la transmission et 0,25 s à la réception du premier caractère)
- unité centrale frontale (de 18 à 50 ms)
- gamme de temps de réponse de l'unité centrale principale (0,5 à 1 ms)
- accès aux disques (0,016 à 2400 ms)

Les trois premiers intervalles tendent à être constants quel que soit le nombre d'utilisateurs reliés au système. On les estime à 500 ms pour le bouclage de ligne et à 50 ms ou moins pour l'unité centrale principale, selon le système actuel.

L'accès aux disques ont un facteur important de ralentissement du système, le temps de réponse étant étroitement lié au nombre d'utilisateurs qui essaient d'avoir accès au système. On peut considérer le disque comme une file d'attente pouvant desservir un seul client à la fois, à un rythme constant de 80 ms par client. La théorie de la file d'attente peut alors servir à calculer le temps d'attente moyen des clients et la variation qui s'applique à cette moyenne. Le graphique ci-joint est le résultat de ces calculs lorsque les demandes d'affichage de pages sont réparties uniformément.

En fait, la répartition des demandes peut être inégale car la consultation des pages d'index de la structure ramifiée peut être très rapide et la lecture des documents très lente. En outre, plusieurs facteurs, dont l'ordre des instructions et la position d'une page sur le disque, font varier le temps de réponse. Cependant, seul un modèle de simulation complexe ou des mesures relevées directement sur le système en opération, pourront déterminer les temps de réponse réeels.

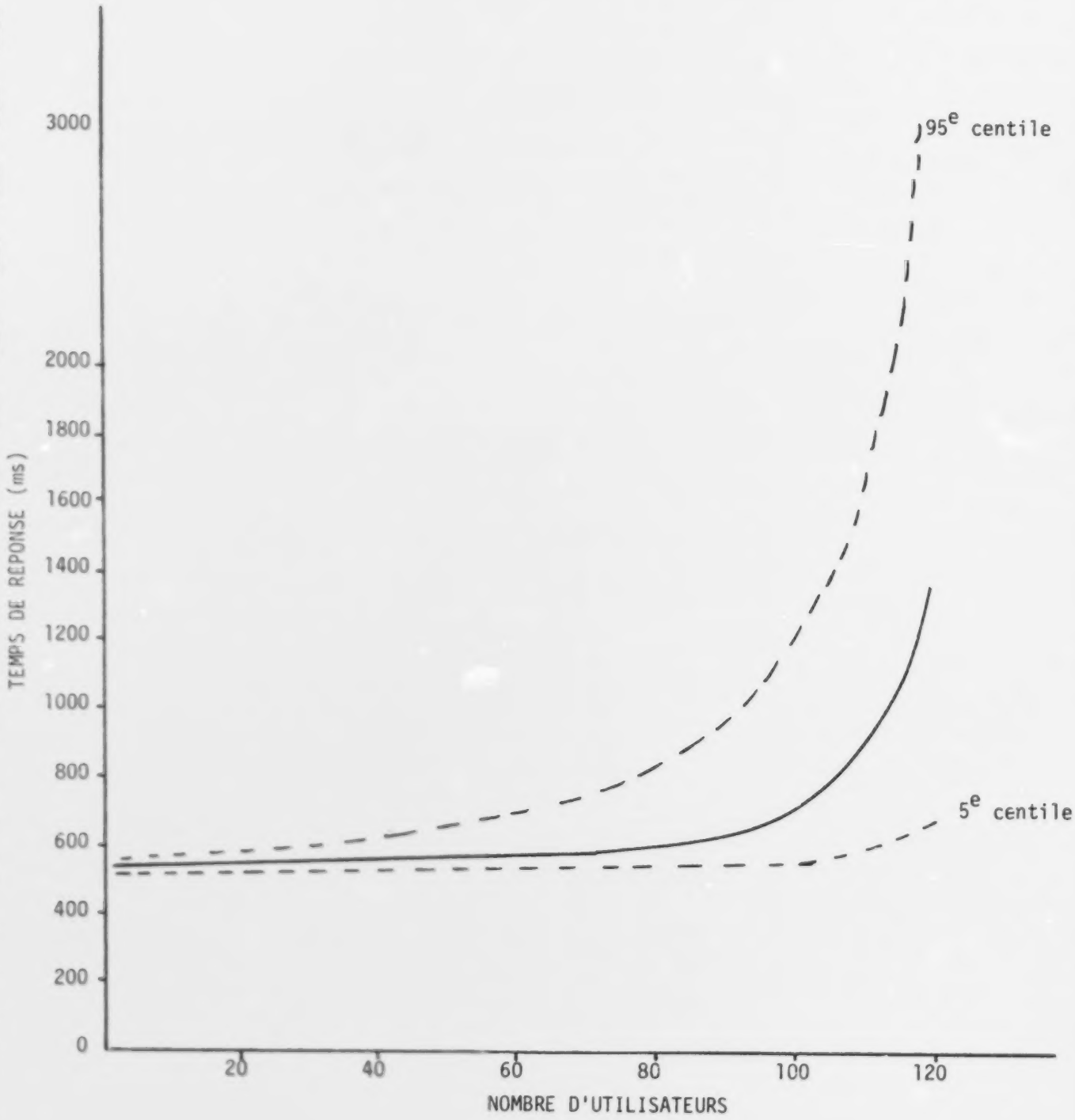
J'espère que ces précisions vous seront utiles. Je vais bientôt rédiger un rapport plus circonstancié sur les temps de réponse et vous en ferai sans retard parvenir un exemplaire.

Robert Baser
Transmission des images

RB: pt

Pièces jointes

BASE DE DONNEE TELEDON



LE TELIDON EN MODE DE RADIODIFFUSION: NOMBRE OPTIMAL DE RUBRIQUES PAR PAGE D'INDEX

Eric Lee

Ministère des Communications

Ottawa (Ontario)

Décembre 1979

LE TELIDON EN MODE DE RADIODIFFUSION: NOMBRE OPTIMAL DE RUBRIQUES PAR PAGE D'INDEX

ERIC LEE*

RESUME

Les conclusions de la présente étude sont les suivantes:

A. LE TELIDON EN MODE DE RADIODIFFUSION AVEC TRANSMISSION PENDANT LES INTERVALLES DE SUPPRESSION VERTICALE

1) DANS LE CAS DES BASES DE DONNEES NE CONTENANT QU'UNE CENTAINE DE PAGES DE DOCUMENTATION, LES FOURNISSEURS D'INFORMATION DEVRAIENT INCLURE DE 7 A 13 RUBRIQUES PAR PAGE. DANS LE CAS DES BASES DE DONNEES CONTENANT ENVIRON 300 PAGES DE DOCUMENTATION, ON DEVRAIT OFFRIR UN CHOIX DE 15 A 20 RUBRIQUES PAR PAGE D'INDEX.

2) L'EMPLOI D'UN TRES PETIT NOMBRE DE RUBRIQUES PAR PAGE (2 OU 3, PAR EXEMPLE) EST NETTEMENT A DECONSEILLER.

3) SI LE NOMBRE MAXIMAL DE RUBRIQUES PAR PAGE EST FIXE A 10, (A) ON DEVRAIT AFFICHER LE PLUS GRAND NOMBRE POSSIBLE DE RUBRIQUES SUR CHAQUE PAGE ET (B) ON DEVRAIT AFFICHER AU MOINS 6 RUBRIQUES PAR PAGE POUR REDUIRE LE PLUS POSSIBLE LE TEMPS DE RECHERCHE.

4) ON DEVRAIT PROBABLEMENT RESTREINDRE A UN MAXIMUM DE 300 LE NOMBRE DE PAGES DES BASES DE DONNEES.

B. LE TELIDON EN MODE DE RADIODIFFUSION AVEC TRANSMISSION SUR LA TOTALITE D'UN CANAL

5) ON PEUT LIMITER A UN MAXIMUM DE 10 LE NOMBRE DE RUBRIQUES PAR PAGE SANS AUGMENTER SENSIBLEMENT LES TEMPS DE RECHERCHE.

6) CHAQUE PAGE DEVRAIT PROPOSER LE PLUS GRAND NOMBRE POSSIBLE DE RUBRIQUES. ON DEVRAIT AFFICHER AU MOINS 6 RUBRIQUES PAR PAGE.

7) MEME SI LE NOMBRE DE PAGES D'UNE BASE DE DONNEES ATTEINT 15 000, UN UTILISATEUR DEVRAIT POUVOIR EN CONSULTER L'INDEX ET REPERER L'INFORMATION QU'IL DESIRE EN MOINS DE 2 MINUTES EN MOYENNE.

* ERIC LEE, Groupe de recherche en comportement
Ministère des Communications, Ottawa (Ontario)

Le présent document vise à déterminer le nombre de rubriques que devrait comporter chaque page d'index de façon à réduire au minimum le temps qu'exige le repérage d'un renseignement (temps de recherche) par l'utilisateur du Télidon en mode de radiodiffusion. Deux méthodes de transmission sont examinées: a) la transmission pendant les intervalles de suppression verticale (ISV), c'est-à-dire la transmission des pages du Télidon pendant les retours de lignes de chaque image (sur une partie seulement d'un canal de télévision) et b) la transmission sur la totalité d'un canal. Trois analyses sont faites:

A) le nombre optimal de rubriques par page en mode de transmission ISV lorsque le nombre de documents de la base de données est invariable et connu,

B) le nombre optimal de rubriques par page en mode de transmission ISV lorsque la dimension de la base de données (pages de documentation et pages d'index) est invariable et connue, et

C) le nombre optimal de rubriques par page en mode de transmission sur la totalité d'un canal lorsque la dimension de la base de données est invariable et connue.

Le temps qu'exige le repérage d'un renseignement dans la base de données (temps de recherche) sera probablement un facteur déterminant de l'attrait du système et de la satisfaction des utilisateurs. Il est permis de présumer que les gens refuseraient d'utiliser un système informatisé de documentation qui leur imposerait des temps de recherche (ST) d'une longueur excessive. Plus on écourtera le temps de recherche moyen, plus le système devrait être bien accueilli et jugé satisfaisant par les utilisateurs et plus les gens devraient être portés à recourir aux services du Télidon.

Le temps de recherche varie en fonction de six facteurs différents: a) le temps que prend la lecture des rubriques de chaque page d'index; b) le temps requis pour repérer et enfoncer les touches sur le clavier; c) le temps moyen d'attente jusqu'à ce que le terminal saisisse la page voulue dans le flot continu de pages de la base de données qui sont diffusées séquentiellement sur le canal de télévision; d) le nombre moyen de pages par document; e) la méthode de recherche des utilisateurs et f) le nombre de rubriques par page d'index.

La relation entre le temps de recherche (ST) et ces six facteurs a été représentée par des formules dont on s'est ensuite servi pour prédire des temps de recherche (ST) selon diverses valeurs estimatives attribuées aux facteurs vitesse de lecture, temps de réponse de l'ordinateur, dimension de la base de données, dimension moyenne des documents, nombre de rubriques par page, etc.

Les trois analyses ont donné les résultats ci-après:

A. Le Télidon en mode de radiodiffusion avec transmission pendant les intervalles de suppression verticale:

1) Lorsqu'il y a trop ou trop peu de rubriques par page, la durée du temps de recherche est excessive. Le temps de recherche le plus court correspond à un nombre intermédiaire de rubriques par page. Voir les figures 2 et 3.

2) Le temps nécessaire pour repérer l'information dans la base de données lorsqu'il n'y a que deux rubriques par page d'index peut être de 2,5 à 6,5 fois plus long que le temps optimal ! (Analyse A).

3) Le nombre optimal de rubriques par page d'index est toujours de 10 rubriques ou plus pour les bases de données contenant 100 pages de documentation ou plus. Voir le tableau 1. (Autrement dit, c'est lorsqu'il y a plus de dix rubriques par page qu'on trouve l'information le plus rapidement.

4) Le temps qu'on mettra à trouver l'information dans la base de données sera supérieur au temps de recherche optimal si on limite à 10 le nombre de rubriques par page. On peut cependant, tout en conservant un maximum de 10 rubriques par page, obtenir des temps de recherche dont l'exédent sur le temps optimal est inférieur à 41%, pourvu que le nombre de rubriques par page soit le nombre optimal "local" (c'est-à-dire le nombre de rubriques par page compris entre 2 et 10 qui réduit le plus possible le temps de recherche).

5) Si le nombre de rubriques par page est limité à 10, on devrait afficher le plus grand nombre possible de rubriques sur chaque page. Il devrait toujours y avoir au moins 6 rubriques par page, sans quoi les temps de recherche deviennent beaucoup trop longs. Lorsque l'affichage est limité

à 10 rubriques par page, le temps de recherche est réduit au minimum dans la plupart des conditions si chaque page contient ce nombre maximal de 10 rubriques.

6) C'est la dimension de la base de données qui détermine, plus que tout autre facteur, le nombre optimal de rubriques par page. Voir les tableaux 1 et 2. Le nombre optimal augmente en fonction de l'augmentation du nombre de pages de documentation (ou du nombre total de pages de la base de données). Voir la figure 1.

7) Le temps qu'exigeront le repérage et l'extraction de l'information dans les bases de données contenant plus de 300 pages de documentation sera probablement supérieur au temps que la plupart des utilisateurs seraient disposés à y consacrer. Les temps de recherche varieraient probablement, en moyenne, de 3 à 7 minutes! (Voir figures 2 et 3.)

B. Le Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal

8) Le nombre optimal de rubriques par index est inférieur ou supérieur à 10 selon les conditions (voir tableaux 3 et 4).

9) Si on limite à 10 le nombre de rubriques par page, le temps de recherche peut être de 21% supérieur au temps optimal. Cependant, dans la plupart des conditions analysées, les temps de recherche se situent à 10% ou moins du temps optimal pour un nombre de 10 rubriques par page.

10) Lorsqu'il y a moins de 6 rubriques par page, les temps de recherche excèdent de beaucoup le temps optimal. Ainsi, à 2 rubriques par page, les temps de recherche sont de 30 à 300% supérieurs au temps optimal. L'augmentation du nombre de rubriques par page peut réduire sensiblement le temps de recherche mais, même à 5 rubriques par page, le temps de recherche peut être de 56% supérieur au temps optimal.

11) DIMENSION DE LA BASE DE DONNEES — TEMPS DE RECHERCHE PREVUS*

5000 pages	de 23 à 79 secondes
10 000 pages	de 41 à 113 secondes
15 000 pages	de 59 à 140 secondes

* Pour une vitesse réelle de transmission des données égale à 362 880 multiplets/s.

DEFINITIONS

temps d'attente (t_{rt}) - le temps d'attente entre la demande d'une page et le début de son affichage sur l'écran. Le temps d'attente correspond au temps de réponse de l'ordinateur dans le cas du Télidon interactif et au temps de radiodiffusion dans le cas du Télidon en mode de radiodiffusion.

temps de recherche (ST)-le temps nécessaire pour effectuer une recherche par arborescences et trouver le document voulu; ce temps est calculé à partir du moment où l'utilisateur entreprend sa recherche au moyen du système Télidon jusqu'au moment où il obtient l'affichage du document sur l'écran. Il comprend le temps requis pour retrouver chacun des index de la séquence qui aboutit au document voulu.

NOTE - le temps de recherche et le temps d'attente ne sont pas du tout synonymes. Le temps d'attente est le temps pendant lequel l'utilisateur attend l'extraction d'une seule page, tandis que le temps de recherche est le temps qu'il consacre à extraire plusieurs pages d'index jusqu'à ce qu'il obtienne le document qu'il recherche.

base de données - la base de données comprend toutes les pages d'information qui peuvent être extraites par un utilisateur et affichées sur un terminal Télidon (ou un téléviseur modifié). Elle comprend à la fois les pages de documentation et les pages d'index.

document - c'est un recueil ou un ensemble de pages (dites pages de documentation) contenant l'information qui intéresse l'utilisateur. Une page de documentation est une page de contenu informationnel. Une base de données peut contenir de nombreux documents et chaque document peut comprendre de nombreuses pages de documentation.

index - c'est l'ensemble des pages (dites pages d'index) dont l'utilisateur peut se servir pour repérer des documents. Chaque page d'index propose un certain nombre de rubriques. Une page d'index est également appelée menu.

p = nombre total de pages d'une base de données, y compris les pages d'index et les pages de documentation;

d = nombre de documents dans la base de données;

i = nombre de pages d'index dans la base de données;

a = nombre de rubriques présentées sur chaque page d'index (menu).

INTRODUCTION

Les présentes analyses s'appuient sur une analyse semblable déjà effectuée relativement au Télidon interactif (voir note technique BRG 79-4). Le lecteur est invité à consulter celle-ci pour connaître plus en détail les hypothèses et les conditions de genre d'analyse.

Pour aider l'utilisateur à consulter une base de données, on peut lui fournir un index à structure ramifiée. Lorsqu'il veut repérer un document dans la base de données au moyen de cette structure, l'utilisateur doit choisir parmi les rubriques reproduites sur chacune des pages d'un ensemble de pages d'index. Pour une base de données particulière, le nombre de pages d'index à consulter pour repérer le document recherché dépend du nombre de rubriques figurant sur chaque page d'index et vice versa. L'augmentation du nombre de rubriques figurant sur chaque page se traduit par une diminution du nombre de pages d'index à consulter pour retrouver un document.

Dès lors qu'il y aura moins de pages d'index à consulter pour repérer un document, il y aura aussi diminution du temps total consacré par les utilisateurs à obtenir l'affichage des pages d'index. Du même coup, par contre, en augmentant le nombre de rubriques affichées sur chaque page, on augmente le nombre de rubriques à lire à chaque page et, de ce fait, on augmente le temps qu'il faut à l'utilisateur pour consulter chaque page d'index. L'effet global du nombre de rubriques par page sur le temps de recherche (temps total consacré par un utilisateur au repérage de l'information voulue dans la base de données du système télétexte) dépend des facteurs suivants: a) temps que prend la lecture des rubriques de chaque page d'index; b) temps requis pour repérer et enfoncer les

touches sur le clavier; c) temps moyen d'attente jusqu'à ce que le terminal saisisse la page voulue dans le flot continu de pages de la base de données qui sont diffusées séquentiellement sur le canal de télévision; d) nombre moyen de pages par document et e) méthode de recherche des utilisateurs. L'importance relative de ces cinq facteurs déterminera l'augmentation ou la diminution du temps de recherche total lorsqu'on accroît le nombre de rubriques par page.

Le temps qu'exige le repérage d'un renseignement dans la base de données, c'est-à-dire le temps de recherche, sera probablement un facteur déterminant de l'attrait du système et de la satisfaction des utilisateurs. Il est permis de présumer que les gens refuseront d'utiliser un système informatisé de documentation qui leur imposerait des temps de recherche (ST) d'une longueur excessive. Ainsi, plus on écourtera le temps de recherche moyen, plus le système devrait être bien accueilli et jugé satisfaisant par les utilisateurs et plus les gens devraient être portés à recourir aux services du Télidon.

ANALYSE A

LE TELIDON EN MODE DE RADIODIFFUSION AVEC TRANSMISSION PENDANT LES INTERVALLES DE SUPPRESSION VERTICALE (NOMBRE OPTIMAL DE RUBRIQUES EN FONCTION DU NOMBRE DE DOCUMENTS)

Dans l'analyse du système Télidon interactif (note technique BRG 79-4), la relation entre le nombre de documents de la base de données (d), le nombre de rubriques par page d'index (a) et le nombre de pages d'index qu'il faut consulter pour repérer le document voulu (p) était représentée par:

1) $d = a^p$

ou

2) $p = \log d / \log a$

Lorsque p, a et d représentent tous des nombres entiers, l'équation 2 donne le nombre exact de niveaux d'index qu'il faut consulter pour extraire un document. Dans tous ces cas, l'extraction du document exige exactement p pages d'index. Cependant, lorsque l'extraction de certains documents de la base de données exige un nombre plus élevé de pages d'index, le nombre moyen de pages d'index requises pour extraire un document choisi au hasard peut varier du minimum au maximum de pages d'index requises pour consulter un document dans la base de données. Autrement dit,

$$p_n \leq \bar{p} \leq p_m$$

où $\bar{p}$ = durée moyenne de la recherche par arborescence (c'est-à-dire le nombre moyen de pages d'index qu'il faut consulter pour extraire un document donné);

p_m = nombre maximal de pages d'index requises pour extraire un document dans une base de données particulière;

p_n = nombre minimal de pages d'index requises pour extraire un document dans une base de données particulière.

Dans ce cas, l'équation 2 donne une approximation du nombre moyen de pages d'index.

Lorsque la base de données est volumineuse et les temps de réponse de l'ordinateur assez rapides (comme ceux du Télidon dont on fait mention dans la note technique BRG 79-4), l'approximation du nombre moyen de pages d'index obtenu au moyen de l'équation 2 est plus que suffisante pour servir au calcul des temps de recherche. Cependant, si la base de données est plus restreinte et les temps d'attente assez longs, comme ce sera le cas pour le Télidon en mode de radiodiffusion, l'équation 2 peut conduire à des erreurs importantes. L'analyse qui suit servira à calculer avec précision les temps de recherche en tenant compte d'une estimation plus précise du nombre de pages d'index qu'il faut consulter pour extraire un document donné.

Supposons que p_m représente le nombre maximal de pages d'index qu'il faut consulter pour extraire un document d'une base de données particulière. Le nombre maximal de pages d'index au niveau p_m de la base de données correspond alors à $a^{p_m} - 1$. Si chacune de ces pages d'index donnait accès à a documents, le nombre maximal de documents auxquels il serait possible d'avoir accès au plus haut niveau d'index p_m est a^{tpm}. Cependant, si le nombre réel de documents de la base de données est inférieur au nombre maximal possible (c'est-à-dire si $d < a^{tpm}$), il n'y aura pas de documents correspondant à $a^{tpm} - d$ des positions possibles au niveau inférieur de la structure arborescente et l'on pourra donc éliminer de la structure ces positions inoccupées. En outre, certaines pages d'index du niveau p_m seront inutiles (puisqu'il n'y a aucun document à retrouver). Chacune de ces pages d'index inutiles au niveau p_m peut être remplacée par l'une des pages de documentation d du niveau $p_m + 1$ de la structure, réduisant ainsi de a le nombre de documents accessibles au niveau $p_m + 1$ de la structure. Cependant, comme on a fait gagner un niveau à un document du niveau inférieur ($p_m + 1$) de la structure, le nombre de positions de documents qu'il est inutile de consulter (parce que ces documents n'existent pas) est réduit de $a - 1$ seulement. Autrement dit, on peut considérer que les pages d'index inutiles au niveau p_m

de la structure donnent accès à un document réel et à $\underline{a} - 1$ documents inexistants. Les documents inexistants disparaissent et, comme il ne reste qu'un document réel, celui-ci peut prendre la place de la page d'index inutile au niveau $\underline{p}_m$. Ainsi, chaque fois qu'on éléminera $\underline{a} - 1$ documents au niveau $\underline{p}_m + 1$ de la structure, on placera un document réel au niveau $\underline{p}_m$ de la structure. Le nombre de documents accessibles à partir de l'avant-dernier ($\underline{p}_m - 1$) et du dernier niveau ($\underline{p}_m$) d'index de la structure sont, respectivement:

3) $$n = \left\lfloor \frac{a^{pm} - d}{a - 1} \right\rfloor$$

4) $$m = d - n$$

où n = nombre de documents dont l'extraction exige $\underline{p}_m - 1$ pages d'index par document (autrement dit, le nombre de documents au niveau $\underline{p}_m$ de la structure);

m = nombre de documents dont l'extraction exige $\underline{p}_m$ pages d'index par document (autrement dit, nombre de documents au niveau $p_m - 1$ de la structure);

p_m = nombre maximal de pages d'index qu'il faut consulter pour extraire un document;

$\lfloor\ \rfloor$ = la valeur du terme compris à l'intérieur des crochets partiels doit être tronquée. Autrement dit, on doit en faire un nombre entier, en laissant tomber les décimales. (Par exemple, si $a = 3$, $d = 12$ et $p_m = 3$, alors $n = \lfloor 7,5 \rfloor = 7$.)

A noter que tous les documents $\underline{d}$ de la base de données sont accessibles à partir des deux derniers niveaux de la structure.

Le terme figurant à l'intérieur des crochets partiels de l'équation 3 peut être un nombre fractionnaire. Cependant, comme le nombre de documents au niveau p_m de la structure doit être un nombre entier, les crochets partiels sont nécessaires pour indiquer qu'il faut rendre le terme entier. Ce système ne permet pas d'avoir des fractions de documents. Le terme de droite est tronqué plutôt qu'arrondi à l'entier supérieur parce que, lorsque a-1 ne se divise pas également en $a^{tpm} - d$, il reste moins de a-1 documents inutiles à la base de la structure. Comme la possibilité d'afficher un ensemble complet de a-1 documents disparaît chaque fois qu'on met un document existant à la place d'une page d'index au niveau p_m de la structure, il devient impossible d'afficher certains documents existants parce qu'il n'y a plus que a-1 positions de documents inutiles. Il faut donc afficher les documents existants supplémentaires à la base de la structure (c'est-à-dire au niveau $p_m - 1$).

La valeur de p_m de l'équation 3 est obtenue par l'équation suivante:

5) $p_m = \lceil \log d / \log a \rceil$

où les crochets partiels indiquent que la valeur du terme qu'ils renferment doit être arrondie à l'entier supérieur. (Par exemple, si $d = 12$ et $a = 3$, $p_m = \lceil 2{,}26 \rceil = 3$.)

Le terme de droite de l'équation 5 est arrondi à l'entier supérieur parce que l'obtention d'une fraction (si petite soit-elle) signifie nécessairement qu'il faut ajouter à un endroit quelconque de la structure un autre niveau pour permettre l'affichage des documents excédentaires. Prenons, par exemple, une base de données contenant 100 documents et offrant un choix de 10 rubriques par page d'index ($d = 100$; $a = 10$). Deux niveaux d'index seulement suffisent pour avoir accès à n'importe quel des 100 documents. Mais l'addition d'un seul autre document à la base de données exige l'addition d'un troisième niveau auxpages d'index pour donner accès à ce document supplémentaire.

Il ne faut pas oublier que chaque document est accessible à partir du dernier ou de l'avant-dernier niveau de la structure.

Si l'on suppose que chaque document est consulté un nombre égal de fois, la probabilité que le document recherché exigera $\underline{p}_m$ niveaux d'index m/d parce que $\underline{m}$ des $\underline{d}$ documents sont au niveau $\underline{p}_m$. De même, la probabilité que le document recherché exigera $\underline{p}_m$ - 1 niveaux d'index est n/d:

6) $P(I = p_m) = m/d$

7) $P(I = p_{m_0} 1) = n/d$

où I = nombre de niveaux d'index qu'il faut consulter;
$P(I = p_m)$ = probabilité que l'extraction d'un document exigera $\underline{p}_m$ niveaux d'index.

Par conséquent, le temps moyen de la recherche par arborescences (nombre moyen de pages d'index à consulter pour repérer l'information voulue) est donné par l'équation suivante:

8) $\overline{p} = (n/d)\ (p_m - 1) + (m/d)\ (p_m)$

$= p_m - n/d$

John Storey estime qu'il est possible de faire défiler 100 pages d'une base de données (un cycle complet) en 20 secondes en utilisant les lignes invisibles du faisceau électronique sur un canal de télévision donné. Autrement dit, il faut un cinquième de seconde pour transmettre chaque page d'information au moyen du système Télidon en mode de radiodiffusion. Étant donné, cependant, que la demande d'information peut survenir à n'importe quel moment de ce cycle de 20 secondes, le nombre prévu de pages (d'index et de documentation) qui devront défiler avant que le terminal ne saisisse la page voulue sera ici 1/2 P. Par conséquent:

9) $t_{rt} = t_b \cdot 1/2\ P$

$= 1/5 \cdot 1/2\ P$

$= 1/10\ P$

où t_{rt} = temps de réponse prévu pour que la page recherchée soit affichée en mode de radiodiffusion (c'est le temps d'attente);

t_b = temps moyen de diffusion d'une seule page d'information;

P = nombre total de pages de la base de données (pages de documentation et pages d'index).

10) $t_c = t_b \cdot P$

$= 1/5\ P$

où t_c = temps d'attente maximal jusqu'à ce que la page recherchée soit affichée en mode de radiodiffusion. Ce temps est également appelé durée du cycle.

Le nombre total de pages de la base de données (P) qui comprend à la fois les pages d'index et les pages de documentation est donné par l'équation suivante:

11) $P = d + i$

où i - nombre de pages d'index. Le k^e niveau de l'index se compose de a^{k-1} pages. Le nombre total de pages d'index est la somme du nombre de pages d'index à chaque niveau de la structure ramifiée.

12) $i = a^0 + a^1 + \ldots + a^{p_m - 2} + (a^{p_m - 1} - n)$

Le dernier terme de l'équation 12 représente le nombre de pages d'index au dernier niveau d'index (c'est-à-dire p_m). On ne doit pas oublier qu'on a remplacé n du nombre maximal possible de pages d'index à ce niveau par des pages de documentation. La somme de tous les termes précédents donne:

13) $i = \sum_{k=0}^{p_m - 1} a^k + (a^{p_m - 1} - n)$

Par substitution, l'équation 11 équivaut à:

14) $P - d + \sum_{k=0}^{p_m - 1} a^k + a^{p_m - 1} - n)$

15) $= d + \frac{a^{p_m - 1} - 1}{a - 1} + a^{p_m - 1} - n$

Le nombre de pages d'index, $\underline{i}$, est égal au nombre de pages d'index que comprennent les niveaux 1 à $p_m - 1$ de la base de données plus le nombre de pages d'index au niveau $\underline{p_m}$ de la base de données (ne pas oublier que $\underline{n}$ pages du niveau $\underline{p_m}$ contiendront des documents).

Le temps de réponse de l'ordinateur est donc:

$$16) \quad t_{rt} = 1/10 \left(d + \frac{a^{p_m - 1} - 1}{a - 1} + a^{p_m - 1} - n\right)$$

La valeur de ST, qui représente le temps de recherche prévu d'un document par le système télétexte, est donnée par l'équation ci-dessous qui correspond à l'équation 4 de la note technique BEG 79-4 (sauf que $\overline{p}$ remplace $\underline{p}$):

$$17) \quad ST = \overline{p} \left(E(A) \cdot t_a + t_p\right) + t_p$$

qui, en remplaçant t_p par $t_{rt} + t_{kp}$, équivaut à

$$18) \quad ST = \overline{p} \left(E(A) \cdot t_a + t_{kp} + t_{rt}\right) + t_{kp} + t_{rt}$$

où ST = temps de recherche total prévu:

$\overline{p}$ = nombre moyen de pages d'index qu'il faut consulter pour extraire un document;

E(A) = nombre prévu du rubriques qu'il faut lire sur une page avant de repérer celle qu'on cherche;

t_a = temps requis pour lire une rubrique (vitesse de lecture de l'utilisateur);

t_p = temps requis pour enfoncer les touches et déclencher l'affichage de la page suivante (temps de réponse du Télidon en mode de radiodiffusion);

t_{rt} = temps de réponse prévu pour que la page recherchée soit affichée en mode de radiodiffusion;

t_{kp} = temps requis pour enfoncer une touche.

On remarquera que les deux derniers termes ($t_{kp} + t_{rt}$) de l'équation 18 représentent le temps nécessaire pour extraire la page de documentation, tandis que le premier terme de droite de l'équation représente le temps requis pour extraire les $\overline{p}$ pages d'index.

Dans le cas de toutes les analyses citées dans cette note technique, l'équation 18 a servi à faire des prédictions et on a toujours supposé qu'il fallait une seconde pour enfoncer les touches appropriées du clavier ($t_{kp} = 1$). D'après nos observations, il semble qu'il faille environ une seconde pour enfoncer les touches après qu'une rubrique a été choisie dans une page d'index. En outre, dans tous les cas, nous supposons qu'il faut un cinquième de seconde pour diffuser chacune des pages de la base de données ($t_b = 0,2$ s).

Exemple: Combien de temps prend, en moyenne, une recherche dans une base de données du Télidon en mode de radiodiffusion contenant 100 documents et offrant un choix de 9 rubriques par page d'index si l'utilisateur a recours à la méthode de recherche exhaustive et met un quart de seconde à lire chaque rubrique?

Autrement dit, $d = 100$, $a = 9$, $t_a = 1/4$, $E(A) = a$.

Le nombre maximal de pages d'index consultées est

$$p_m = \lceil \log d / \log a \rceil = \lceil 2/0,95424 \rceil = \lceil 2,0959 \rceil = 3$$

Le nombre de documents dont l'obtention exige l'extraction de 2 pages d'index est

$$n = \left\lfloor \frac{a^{p_m} - d}{a - 1} \right\rfloor = \left\lfloor \frac{9^3 - 100}{9 - 1} \right\rfloor = \lfloor 78,625 \rfloor = 78$$

Le nombre de documents dont l'obtention exige l'extraction de 3 pages d'index est

$$m = d - n = 100 - 78 = 22$$

$$\overline{p} = p_m - n/d = 3 - 0,78 = 2,22$$

$$p = d + \frac{a^{p_m - 1} - 1}{a - 1} + a^{p_m - 1} - n$$

$$= 100 + \frac{9^2 - 1}{8} + 9^2 - 78$$

$$= 100 + 10 + 81 - 78 = 113$$

$$t_p = 1/10P + 1 = 1/10\ (113) + 1 = 12,3$$

$$ST(a=9) = \overline{p}\ (E(A)\ .\ t_a + t_p) + t_p$$

$$= 2,22\ (9\ .\ 1/4 - 12,3) + 12,3$$

$$= 44,60 \text{ secondes}$$

Dans ces conditions, le temps de recherche prévu est en moyenne de 44,6 secondes.

ANALYSE B

LE TELIDON EN MODE DE RADIODIFFUSION AVEC TRANSMISSION PENDANT LES INTERVALLES DE SUPPRESSION VERTICALE (NOMBRE OPTIMAL DE RUBRIQUES EN FONCTION DE LA DIMENSION DE LA BASE DE DONNEES)

Si c'est la dimension de la base de données (P) plutôt que le nombre de documents (d) qui est invariable au point de départ, l'analyse ci-dessous peut être utilisée pour prédire le nombre optimal de rubriques en fonction de P. Tout d'abord, il faut trouver la valeur de d à partir de P.

La valeur de d se trouve à partir de l'équation 15.

$$15)\quad P = d + \frac{a^{p_m-1} - 1}{a - 1} + a^{p_m-1} - n$$

Par substitution de n et transposition des termes, nous obtenons

$$d = P - \left(\frac{a^{p_m-1} - 1}{a - 1}\right) - a^{p_m-1} - \left\lfloor \frac{a^{p_m} - d}{a - 1} \right\rfloor$$

Puisque les 3 premiers termes de droite de l'équation correspondent à des nombres entiers et que le quatrière terme est le seul qu'il faut transformer en nombre entier, l'opération représentée par $\lfloor\ \rfloor$ (arrondir à l'entier inférieur) peut s'appliquer à tous les membres de droite de l'équation.

$$d = \left\lfloor P - \frac{a^{p_m-1} - 1}{a - 1} - a^{p_m-1} + \frac{a^{p_m}}{a - 1} - \frac{d}{a - 1} \right\rfloor$$

$$d + \frac{d}{a-1} = \left[P - \frac{a^{Pm}-1}{a-1} - 1 - a^{Pm-1} + \frac{a^{Pm}}{a-1} \right]$$

$$\frac{(a-1)\,d}{a-1} + d = \left[P - \frac{a^{Pm}-1}{a-1} - 1 - a^{Pm-1} + \frac{a^{Pm}}{a-1} \right]$$

$$ad - d + d = \left[(a-1) \left(P - \frac{a^{Pm}-1}{a-1} - 1 - a^{Pm-1} + \frac{a^{Pm}}{a-1} \right) \right]$$

$$d = \left[\frac{a-1}{a} \left(P - \frac{a^{Pm}-1}{a-1} - 1 - a^{Pm-1} + \frac{a^{Pm}}{a-1} \right) \right]$$

$$d = \left[\left(\frac{a-1}{a} \right) P - \frac{a^{Pm}-1}{a} - 1 - (a-1)\, a^{Pm-2} + a^{Pm-1} \right]$$

$$d = \left[\left(\frac{a-1}{a} \right) P - a^{Pm-2} + 1/a - a^{Pm-1} + a^{Pm-1} + a^{Pm-2} \right]$$

$$d = \left[\left(\frac{a-1}{a} \right) P + \frac{1}{a} \right]$$

19) $$d = \left[aP - \frac{P}{a} + 1 \right]$$

La valeur de d obtenue par l'équation 19 comprend deux variables, a et P, toutes deux connues et invariables pour toute simulation. Cette équation a été utilisée pour trouver d dans les prédictions relatives au Télidon en mode de radiodiffusion avec transmission pendant l'intervalle de suppression verticale.

Par transformation, l'équation 19 peut devenir:

$$(20)\quad d = \left\lfloor P - \frac{1}{a}(P - 1) \right\rfloor$$

L'équation 20 montre que le nombre de documents égale le nombre total de pages de la base de données moins le nombre de pages d'index. Autrement dit,

$$(21)\quad i = \left\lfloor \frac{1}{a}(P - 1) \right\rfloor$$

Le nombre de pages d'index dépend uniquement du nombre de rubriques par page et de la dimension de la base de données

Le nombre maximal de pages d'index qu'il faut consulter pour extraire un document peut s'obtenir de deux façons (en supposant que chaque document comprend une seule page). Tout d'abord, en posant que $\underline{p}$ est la plus petite valeur pour laquelle $d + i > P$ où $\underline{d}$, $\underline{i}$, $\underline{P}$ et $\underline{p}$ ont tous été transformés en nombres entiers, on obtient $d = a^p$ et $i = \frac{a^p - 1}{a - 1}$.

Deuxièmement, on peut utiliser l'équation suivante:

$$(22)\quad p = \left\lceil \log \left\lfloor P - 1/a(P - 1) \right\rfloor \ / \log a \right\rceil$$

Les équations 19, 22 et 18 ont servi à prédire des temps de recherche.

ANALYSE C

LE TELIDON EN MODE DE RADIODIFFUSION AVEC TRANSMISSION SUR LA TOTALITÉ D'UN CANAL (NOMBRE OPTIMAL DE RUBRIQUES EN FONCTION DE LA DIMENSION DE LA BASE DE DONNEES)

Lorsqu'un canal complet de télévision est utilisé pour la radiodiffusion d'une base de données, on peut transmettre beaucoup plus de pages d'information en un temps donné que si on n'utilise qu'une partie d'un canal pendant l'intervalle de suppression verticale (voir document ci-joint). Avec une seule modification, l'analyse de la transmission ISV, peut aussi être utilisée pour prédire les temps de recherche (ST) de la transmission sur la totalité d'un canal. Dans le cas de l'analyse de la transmission ISV, on supposait de façon implicite que chaque document contenait exactement une seule page d'information. Pour les bases de données restreintes qui seront probablement transmises pendant les intervalles de suppression verticale, il ne fait aucun doute que l'hypothèse se révélera assez juste. Pour ce qui est des bases de données plus volumineuses qui seront probablement transmises sur la totalité d'un canal, chaque document contiendra probablement de 2 à 5 pages d'information.

L'analyse de la transmission ISV peut aussi être appliquée à la transmission sur la totalité d'un canal si l'on y introduit une nouvelle variable, f, c'est-à-dire le nombre moyen d'images ou de pages d'information d'un document.

Si l'on suppose que chaque document d'une base de données contient f images (pages), l'analyse précédente peut être généralisée par les équations suivantes. Tout d'abord, le nombre total de pages d'une base de données égale le nombre de documents multiplié par le nombre moyen de pages par document plus le nombre de pages d'index. (Evidemment, lorsque f = 1, P = d + i comme l'équation 11.)

22) $P = df - i$

$df = P - i$

où f = nombre moyen de pages ou d'images d'information d'un document.

En remplaçant i par ses équivalents, nous trouvons (voir équation 15):

$$df = \left\lfloor P - \frac{a^{P_m-1} - 1}{a - 1} - a^{P_m-1} + \frac{a^{P_m}}{a - 1} - \frac{d}{a - 1} \right\rfloor$$

$$df + \frac{d}{a - 1} = \left\lfloor P - \frac{a^{P_m-1} - 1}{a - 1} - a^{P_m-1} + \frac{a^{P_m}}{a - 1} \right\rfloor$$

$$\frac{(a - 1)\ df + d}{a - 1} = \left\lfloor P - \frac{a^{P_m-1} - 1}{a - 1} - a^{P_m-1} + \frac{a^{P_m}}{a - 1} \right\rfloor$$

$$adf - df + d = \left\lfloor aP - P - a^{P_m-1} + 1 - a^{P_m} + a^{P_m-1} + a^{P_m} \right\rfloor$$

$$d = 1/(af - f + 1) \left\lfloor aP - P + 1 \right\rfloor$$

24) $$d = \left\lfloor \frac{aP - P + 1}{af - f + 1} \right\rfloor$$

On constate, qu'en posant f = 1 dans la dernière équation (en supposant que la longueur de chaque document correspond exactement à celle d'une page), on résout le cas spécial de d dans l'équation 19.

De façon générale, lorsque chaque document contient en moyenne $\underline{f}$ pages, la valeur de $\underline{p}$ s'obtient par l'équation:

$$25)\quad p = \left\lceil \log \left\lfloor \frac{aP - P + 1}{af - f + 1} \right\rfloor / \log a \right\rceil$$

Après avoir trouvé les valeurs appropriées de $\underline{d}$ et de $\underline{p}$ en utilisant respectivement les équations 24 et 25, l'équation 18 peut être utilisée pour prédire les temps de recherche de la transmission sur la totalité d'un canal.

Le terme t_{rt}, qui représente le temps d'attente jusqu'à ce que la page suivante soit affichée, est égal à la moitié de la durée du cycle $\underline{t_c}$. La durée du cycle est le temps nécessaire pour diffuser une fois chaque page de la base de données.

$$26)\quad t_{rt} = t_c/2$$

La durée du cycle est obtenue par l'équation:

$$27)\quad t_c = BP \cdot P/EDR$$

et le débit réel de transmission des données est

$$28)\quad EDR = BL \cdot NL \cdot C$$

où BP = nombre moyen de multiplets d'information par page du système Télidon (multiplets par page);

P = dimension de la nase de données, c'est-à-dire nombre de pages de la base de données;

EDR = débit réel de transmission des données (multiplets/s), c'est-à-dire vitesse moyenne à laquelle l'information est diffusée;

BL = nombre de multiplets de données (information) par ligne de balayage du faisceau électronique (multiplets/ligne de balayage);

NL = nombre de lignes utiles du balayage par trame (lignes de balayage/trame);

C = vitesse du cycle des trames (trames/cycle).

Ainsi, la durée du cycle est égale au nombre total de lignes véhiculant l'information de la base de données, divisé par la vitesse moyenne de transmission. Le débit réel de transmission des données dépend du nombre de multiplets/ligne de balayage, BL. Nous avons tenu compte de deux débits de transmission BL: 16 et 24 multiplets/ligne de balayage. Les débits réels de transmission correspondants sont 241 920 et 362 880 multiplets/s respectivement, en supposant qu'une page du système Télidon contient, en moyenne, 600 multiplets environ. En outre, soixante trames sont transmises chaque seconde (c'est-à-dire C = 60 trames/s).

HYPOTHESES RETENUES POUR LES SIMULATIONS

Des temps de recherche ont été prévus pour chaque analyse en tenant compte des hypothèses et conditons suivantes.

Analyses A et B

1. Les utilisateurs ont recours à une méthode de recherche exhaustive ou à une méthode de recherche autocessante.

2. Il faut aux utilisateurs de 0,25 à 2 secondes pour lire chaque rubrique.

3. Le nombre de documents varie de 100 à 900 (voir tableau 1) OU le nombre de pages de la base de données varie de 100 à 900 (voir tableau 2).

4. Chaque document contient une seule page d'information.

5. Le débit réel de transmission des données est 3000 multiplets/pour la transmission pendant les intervalles de suppression verticale.

6. Dans l'établissement de la structure ramifiée, chaque page d'index de la base de données offre un choix de a rubriques.

7. Les utilisateurs ne comettent aucune erreur lorsqu'ils effectuent une recherche dans la base de données du Télidon. (Des erreurs se traduiraient par des temps de recherche plus longs que prévus.)

Analyse C

1. Les utilisateurs ont recours à une méthode de recherche exhaustive ou à une méthode de recherche autocessante.

2. Il faut aux utilisateurs de 0,25 à 2 secondes pour lire chaque rubrique.

3. La dimension de la base de données (y compris les pages d'index et les pages de documentation) varie de 5000 à 15 000 pages.

4. Chaque document contient, en moyenne, de 2 à 5 pages d'information.

5. Le débit réel de transmission des données sur la totalité d'un canal correspond soit à 362 880 multiplets/s (24 multiplets par ligne de balayage; 252 lignes par trame; 60 trames/s) ou à 241 920 multiplets/s (16 multiplets par ligne de balayage).

6. Dans l'établissement de la structure ramifiée, chaque page d'index offre un choix de a rubriques.

7. Les utilisateurs ne commettent aucune erreur lorsqu'ils effectuent une recherche dans la base de données du Télidon.

RESULTATS

On ne savait pas exactement combien on pouvait afficher de rubriques sur une page d'index sans diminuer la capacité de lecture de l'utilisateur et sans abréger chaque rubrique au point d'en rendre le sens ambigu. Des exemples de pages d'index comprenant 20, 40, 60 et 80 rubriques sont présentés aux figures 4 a, b, c et d. Comme de nombreuses rubriques de la base actuelle de données du Télidon excèdent la longueur maximale permise pour l'affichage de 60 et 80 rubriques par page, aucun fournisseur d'information n'aura probablement recours à plus de 40 rubriques par page. Par conséquent, nous avons fait varier de 2 à 40 le nombre de rubriques par page pour les prédictions faites à partir des trois analyses qui suivent.

Analyse A

Pour l'établissement des prédictions présentées au tableau 1 et aux figures 1 à 3, nous avons pris systématiquement des bases de données contenant 100, 300, 500 et 900 pages de documentation. Dans tous les cas, la base de données doit être plus grande que le nombre de documents, parce qu'elle comprend à la fois les pages d'index et les pages de documentation. Ainsi, lorsque le nombre de documents est égal à 100 et le nombre de rubriques par page à 10, le nombre de pages de la base de données est égal à 111 ($P = d + i$).

Le nombre optimal de rubriques pour différents nombres de documents, méthodes de recherche et vitesses de lecture est indiqué au tableau 1. Ce tableau montre que, dans les conditions considérées, le nombre optimal de rubriques par page d'index est toujours de 10 ou plus. Ce résultat est en contraste marqué avec celui de l'analyse portant sur le Télidon interactif, où le nombre optimal de rubriques par page était toujours de 10 ou moins. Cette différence entre le nombre optimal de pages du système interactif et celui du système Télidon en mode de radiodiffusion est attribuable aux facteurs suivants: a) les temps de réponse du système Télidon en mode de radiodiffusion sont beaucoup plus longs (de 10 à 90 seconder) que ceux du système Télidon en mode interactif (de 0,5 à 1,5 seconde); b) les bases de données du système Télidon en mode de radiodiffusion sont beaucoup plus restreintes.

Les temps de recherche prévus en proportion du nombre de documents sont présentés aux figures 2 et 3 (pour une transmission ISV). On constate que les temps de recherche les plus courts correspondent à des valeurs intermédiaires de a, c'est-à-dire le nombre de rubriques par page. Lorsqu'il y a trop ou trop peu de rubriques par page, les temps de recherche sont beaucoup plus longs. On obtient les temps de recherche les plus longs lorsqu'il y a seulement 2 rubriques par page d'index. Avec 2 rubriques par page, la longueur des temps de recherche est de 2,5 à 6,5 fois celle du temps optimal!

Le nombre optimal de rubriques par page est toujours de 10 ou plus pour les bases de données contenant 100 pages de documentation ou plus (voir tableau 1). Cependant, contrairement au système Télidon interactif, le nombre optimal dépend ici du nombre de documents de la base de données et il est directement proportionnel au nombre de pages de documentation de la base de données (voir figure 1).

Si le nombre de rubriques par page est limité à 10, le temps de recherche prévu sera plus long que temps optimal. Pourvu qu'on s'en tienne au nombre optimal de 2 à 10 rubriques par page, les temps de recherche ne dépasseront pas de plus de 41% le nombre optimal général. Dans la plupart des cas, le nombre optimal local correspond en fait à 10 rubriques par page. Si chaque page contient moins de 10 rubriques, les temps de recherche seront beaucoup plus longs que le temps optimal. Ainsi, chaque page d'index devrait comprendre le plus grand nombre possible de rubriques. Autrement dit, le nombre de rubriques affichées sur chaque page devrait se rapprocher le plus possible de 10.

Qu'il y ait ou non restriction du nombre de rubriques par page, chaque page devrait comprendre au moins 6 rubriques. Les temps de recherche sont démesurément longs lorsque chaque page contient de 2 à 5 rubriques. Pour 100 pages de documentation, les temps de recherche qu'exigeront les pages de 6 rubriques ou plus n'excéderont pas de plus de 25% le temps optimal. Pour 300 pages de documentation, les temps de recherche qu'exigeront les pages de 6 rubriques ou plus n'excéderont pas de plus de 70% le temps optimal.

Le temps qu'exigeront la recherche et l'extraction de l'information dans les bases de données contenant plus de 300 pages de documentation sera probablement supérieur au temps que la plupart des utilisateurs seraient disposés à y consacrer. Ces temps de recherche varieraient probablement, en moyenne, de 3 à 7 minutes! (Voir figures 2 et 3.) Dans le cas d'une base de données restreinte ne contenant que 100 pages de documentation, les temps de recherche prévus varient de 35 à 80 secondes (pourvu que chaque page contienne de 5 à 10 rubriques); voir figures 2a et 3a. Dans le cas d'une base contenant 300 pages de documentation, les temps de recherche se situeraient entre 100 et 175 secondes (si chaque page contenait de 15 à 20 rubriques); voir figures 2b et 3b.

Analyse B

Les nombres optimaux prévus pour différentes bases de données (transmission ISV) sont présentés au tableau 2. Ces résultats se comparent à ceux de l'analyse A puisque le nombre optimal de rubriques par page est de 10 ou plus dans la plupart des conditions. Le tableau 2 montre que le nombre optimal ne dépend ni de la méthode de recherche, ni de la vitesse de lecture. Cependant, il est directement proportionnel à la dimension de la base de données. Le temps de recherche prévu est représenté graphiquement en fonction du nombre de rubriques par page aux figures 6 et 7.

Analyse C

Les nombres de rubriques optimaux prévus pour diverses bases de données (transmission sur la totalité d'un canal) sont présentés aux tableaux 3 et 4. La méthode de recherche utilisée n'influe pas sensiblement sur le nombre optimal de rubriques par page. Celui-ci, par contre, augmente lorsque diminue le temps de lecture. Le nombre optimal de rubriques par page est directement proportionnel à la dimension de la base de données et inversement proportionnel au nombre moyen de pages par document. Le fait de porter le débit réel de transmission

des données de 241 920 à 362 880 multiplets/s a très peu d'effet sur le nombre optimal de rubriques par page. Celui-ci est tantôt supérieur, tantôt inférieur à 10 rubriques par page d'index. Les figures 5a à 5h présentent les temps de recherche prévus dans le cas de l'utilisateur moyen du service Télidon (c'est-à-dire pour une vitesse de lecture de 1 seconde par rubrique).

Le tableau 5 présente la gamme des temps de recherche escomptés qui devraient correspondre à différentes bases de données et à différents débits de transmission. L'augmentation de la dimension de la base de données semble se traduire par une augmentation linéaire des temps de recherche (ST). Le débit le plus élevé de transmission de données (362 880 multiplets/s réduit le temps de recherche, mais non pas dans la proportion escomptée (pour une augmentation de 50% du débit de transmission, le temps de recherche est réduit de 10 à 25% seulement). Dans la plupart des cas, les temps de recherche (ST) sont de l'ordre de 1 à 2 minutes. Ces temps de recherche semblent donc acceptables, mais seule une recherche empirique permettra d'en confirmer l'exactitude.

Lorsque chaque page contient moins de 6 rubriques, la longueur des temps de recherche est excessive par rapport au temps optimal. Ainsi, à 2 rubriques par page, les temps de recherche sont de 30 à 300% plus longs que le temps optimal. On peut réduire considérablement le temps de recherche en augmentant le nombre de rubriques par page. Cependant, même à 5 rubriques par page, il dépasse encore de 56% le temps optimal.

L'excédent du temps de recherche, par rapport au temps optimal, peut atteindre 21% si on limite à 10 le choix de rubriques affichées sur chaque page. Cependant, dans la plupart des conditions, les temps de recherche obtenus à 10 rubriques ou moins par page n'excédent le temps optimal que de 10% ou moins.

Les figures 5a à 5h montrent la relation entre le temps de recherche et le nombre de rubriques par page dans le cas d'un lecteur moyen. Les figures 6 à 17 montrent la relation entre le temps de recherche et le nombre de rubriques par page dans le cas des lecteurs les plus lents et des lecteurs les plus rapides. Ainsi, aux figures 6 à 17, les temps de recherche prévus pour tous les lecteurs devraient se situer quelque part entre les deux courbes de fonctions extérieures des graphiques.

DISCUSSION

Un clavier numérique serait fort utile si le nombre optimal de rubriques par page était toujours de 10 rubriques ou moins, car il serait alors possible d'utiliser des chiffres plutôt que des lettres pour coder les pages d'index. Or, comme le nombre optimal est parfois supérieur à 10 rubriques par page pour la transmission sur la totalité d'un canal, un clavier alphabétique semble nécessaire. On peut cependant envisager deux autres solutions d'un égal mérite qui permettraient de conserver le clavier numérique. La première consisterait à assigner à chacune des rubriques un code à deux chiffres (voir figure 4). Lorsque la numérotation d'une page de documentation est déterminée par la position du document dans l'index ramifié, le fait de doubler le nombre de chiffres attribués à chaque rubrique a essentiellement pour effet de doubler le nombre de chiffres nécessaires au codage de chaque page de documentation. Doubler le nombre de chiffres créerait probablement un sérieux problème dans le cas des bases de données très importantes comme celles du système interactif Télidon (par exemple, dans une structure ramifiée comprenant 16 niveaux, on se retrouverait avec un code de 32 chiffres pour chaque page de documentation). Toutefois, on peut présumer que l'effet du doublement ne prendra que des proportions relativement minimes dans le cas des bases de données restreintes comme celles du système Télidon en mode de radiodiffusion avec transmission ISV (de 2 à 4 pages d'index, c'est-à-dire de 4 à 8 chiffres, suffiraient alors pour avoir accès à un document donné). Pour ce qui est des bases de données relativement restreintes qui alimenteront probablement le système Télidon en mode de radiodiffusion avec transmission ISV (bases comprenant de 100 à 1000 pages), le fait de doubler le nombre de chiffres ne risque guère d'incommoder l'utilisateur (puisque de 4 à 8 chiffres seulement lui permettront quand même d'avoir directement accès à un document). De plus, si une base contient moins de 100 documents, le nombre optimal de rubriques par page sera inférieur à 10 dans la plupart des conditions. Dans ce cas, il ne fait aucun doute qu'un clavier numérique donnerait satisfaction.

La deuxième solution qui permettrait de conserver un clavier numérique consiste à limiter à 10 le nombre maximal de rubriques par page. Cette solution est particulièrement attrayante dans le cas de la transmission sur la totalité d'un canal parce que, dans les conditions les moins favorables, les temps de recherche correspondant à ce nombre maximal de 10 rubriques n'excéderaient pas de plus de 21% le temps correspondant au nombre optimal. En outre, dans la plupart des conditions, les temps de recherche n'excéderaient pas de plus de 10% le temps de recherche optimal.

Il est donc fortement recommandé de limiter à 10 le nombre de rubriques par page de façon à pouvoir utiliser un clavier numérique dans le cas du système Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal. Malheureusement, aucune recommandation précise ne peut être faite dans le cas de la transmission ISV. Dans certains cas, les temps de recherche correspondant à un maximum de 10 rubriques par page excèdent modérément (par plus de 50%) le temps optimal. De plus, le nombre optimal est toujours 10 pages ou plus pour toute base de données contenant 100 pages de documentation ou plus. Il n'est

pas facile de choisir entre une méthode de codage par doublement des chiffres et la fixation d'un nombre maximal de rubriques par page. Il peut être nécessaire d'effectuer une recherche empirique pour cerner de plus près les conséquences de ces deux solutions.

Les figures 2 et 3 montrent que, pour une base de données très restreinte contenant environ 100 pages de documentation, l'utilisateur du système Télidon en mode de radiodiffusion ne trouvera sans doute pas que les temps de recherche sont trop longs (pour la plupart des utilisateurs ils seraient probablement de 0,66 à 1,33 minute pourvu que chaque page d'index comprenne le nombre optimal de rubriques). Dans le cas d'une base de données contenant 300 documents, les utilisateurs prendraient probablement de 2 à 3 minutes pour repérer l'information voulue dans la structure ramifiée. Il n'est pas certain que les utilisateurs accepteraient de consacrer plusieurs minutes à la recherche de données. Cependant, dans le cas des deux bases de données les plus importantes (500 et 900 pages de documentation), les temps de recherche optimums sont de l'ordre de 3 à 4,5 minutes et de 5 à 7 minutes! A ce compte, peu de gens accepteraient probablement d'utiliser le Télidon. Une étude empirique est nécessaire pour déterminer exactement combien de temps les utilisateurs sont disposés à consacrer à la recherche.

Contrairement au temps de recherche du système Télidon interactif, le temps de recherche du système Télidon en mode de radiodiffusion (ISV) varie selon la dimension de la base de données. En effet, le temps de réponse du signal radiodiffusé (représenté par t_{rt}) dépend directement de la dimension de la base de données. Dans le cas du système Télidon interactif, le temps de recherche ne varie pas en fonction de la dimension de la base de données parce que le temps de réponse de l'ordinateur n'est pas fonction de la dimension de la base de données.

Government of Canada / Gouvernement du Canada

MEMORANDUM **NOTE DE SERVICE**

TO / A: Eric Lee

FROM / DE: Bob FitzGerald

SECURITY CLASSIFICATION / DE SÉCURITÉ:

OUR FILE / N/RÉFÉRENCE:

YOUR FILE / V/RÉFÉRENCE:

DATE: le 17 septembre 1979

SUBJECT / OBJET: TEMPS D'ATTENTE DU SYSTÈME TÉLIDON EN MODE DE RADIODIFFUSION (AVEC TRANSMISSION SUR LA TOTALITÉ D'UN CANAL)

Pour faire suite à notre conversation téléphonique de ce matin, vous trouverez ci-dessous quelques calculs du temps moyen d'attente relativement à la réception d'une page du système Télidon en mode de radiodiffusion. A noter que le temps d'attente est la somme de l'intervalle entre la demande de l'utilisateur et la réception de la première ligne d'une page, et le temps réel de transmission de cette page.

Hypothèses

- Transmission sur la totalité d'un canal (N_L = 252 lignes utiles du balayage télévisuel)
- Nombre de multiplets de données par ligne du balayage (B_L) = 16, 20 ou 24
- Nombre moyen de multiplets par page du Télidon (B_p) $\simeq$ 600
- Nombre de pages du cycle de diffusion (N_p) $>$ 80 pour qu'il y ait un temps minimal de 5 ms entre deux lignes d'une page donnée.
  Débit réel de transmission des données (f_D) = BL X NL X 60

Exemple BL = 24

$$f_D = 24 \frac{\text{multiplets}}{\text{ligne}} \times 252 \frac{\text{lignes}}{\text{trame}} \times 60 \frac{\text{trames}}{\text{seconde}} = 362\,880 \frac{\text{multiplets}}{\text{seconde}}$$

DURÉE DU CYCLE (t_c) = $\frac{600 \times N_p}{f_D}$

Exemple: N_p = 10 000, $t_c = \frac{600 \times 10\,000}{362\,880} = 16{,}5$ s

ou

$$N_p = \frac{t_c \times f_D}{600}$$

Exemple: t_c = 20 s, $N_p = \frac{20 \times 362\,880}{600}$ = 12 096 pages (en moyenne)

La première composante du temps d'attente t_1 est $0 \leq t_1 \leq t_c$

Valeur moyenne de t_1 = $\frac{t_c}{2}$

Par exemple, si t_c = 20 s, t_1 = 10 s

La deuxième composante du temps d'attente t_2 est le temps requis pour transmettre la page (moyenne) voulue.

Le nombre de lignes du balayage télévisuel requis pour transmettre la page moyenne de 600 multiplets est

$$L_p = \frac{600}{BL}$$

Exemple: BL = 24, LP = $\frac{600}{24}$ = 25 lignes

Si l'on suppose que les lignes d'information sont intercalées de sorte qu'une ligne sur 100 du balayage télévisuel contient des données de la page recherchée, le nombre total de lignes du balayage dans l'intervalle requis pour la transmission d'une page complète est 25 X 100 = 2500 et le temps de transmission est

$$t_2 = \frac{2500 \text{ lignes}}{252 \text{ lignes/trame X } 60 \text{ trames/s}} = 0{,}165 \text{ s}$$

A partir des hypothèses précédentes, on conclut que le temps d'attente total est

$t = t_1 + t_2 = 10{,}165$ s

Précisons que la grande différence entre les valeurs de t_1 et de t_2 est particulièrement marquée dans le cas d'une transmission sur la totalité d'un canal. Les résultats sont bien différents pour une transmission pendant l'intervalle de suppression verticale (ISV).

J'espère que cette analyse vous sera utile. Vu l'urgence de votre demande, je n'ai pu revoir ces calculs avec John Storey et ne peux donc garantir que ma méthode de calcule est exempte d'erreurs ou que mon analyse reflète les vues les plus récentes de John sur le sujet. Il serait peut-être souhaitable que vous en discutiez avec lui à son retour lundi prochain avant de publier votre document.

Bob FitzGerald
Transmission des images

BFG/cb

c.c. J.R. Storey

TABLEAU 1

NOMBRE OPTIMAL DE RUBRIQUES PAR PAGE D'INDEX
DU SYSTÈME TÉLIDON EN MODE DE RADIODIFFUSION
AVEC TRANSMISSION ISV[a] (ANALYSE A)

MÉTHODE DE RECHERCHE DE L'UTILISATEUR	TEMPS DE LECTURE PAR RUBRIQUE	NOMBRE DE PAGES DE DOCUMENTATION DANS LA BASE DE DONNÉES[b]			
		100	300	500	900
EXHAUSTIVE	.25 sec	11 (10-21)	21 (18-27)	26 (23-33)	32 (32-40)
	.50	10 (10-11)	18 (18-20)	23 (23-25)	30 (30-32)
	1.00	10 (10)	18 (17-18)	23 (23)	30 (30)
	2.00	10 (10)	17 (17-18)	22 (22-23)	30 (30)
AUTOCESSANTE	.25 sec	26 (13-40)	31 (22-40)	40 (28-40)	39 (34-40)
	.50	11 (10-21)	21 (18-27)	26 (23-33)	32 (30-40)
	1.00	10 (10-11)	18 (18-20)	23 (23-25)	30 (30-32)
	2.00	10 (10)	18 (17-18)	23 (23)	30 (30)

[a]Contrairement à ce qui se produit dans le cas du système Télidon interactif, le nombre optimal de rubriques par page du système Télidon en mode de radiodiffusion varie selon la dimension de la base de données (c'est-à-dire le nombre total de documents).

[b]Les inscriptions du tableau représentent le nombre optimal de rubriques pour les conditions précisées. Nous donnons entre parenthèses, sous chaque nombre optimal, les nombres minimal et maximal de rubriques pour lesquels le temps de recherche total n'excède pas de plus d'une seconde, le temps optimal. Ainsi, lorsque le nombre optimal est de 11 rubriques par page, on peut afficher de 10 à 21 rubriques par page sans que le temps de recherche augmente de plus d'une seconde.

TABLEAU 2

NOMBRE OPTIMAL DE RUBRIQUES PAR PAGE D'INDEX
DU SYSTÈME TÉLIDON EN MODE DE RADIODIFFUSION
AVEC TRANSMISSION ISV (ANALYSE B)

MÉTHODE DE RECHERCHE DE L'UTILISATEUR	TEMPS DE LECTURE PAR RUBRIQUE	NOMBRE DE PAGES DANS LA BASE DE DONNEES			
		100	300	500	900
EXHAUSTIVE	.25 sec	10	17	22	30
	.50	10	17	22	30
	1.00	10	17	22	30
	2.00	9	17	22	30
AUTOCESSANTE	.25 sec	12	17	22	30
	.50	10	17	22	30
	1.00	10	17	22	30
	2.00	10	17	22	30

TABLEAU 3

NOMBRE OPTIMAL DE RUBRIQUES PAR PAGE D'INDEX
DU SYSTEME TÉLIDON EN MODE DE RADIODIFFUSION AVEC TRANSMISSION
SUR LA TOTALITÉ D'UN CANAL A 241 920 MULTIPLETS PAR SECONDE[a]
(ANALYSE C)

MÉTHODE DE RECHERCHE DE L'UTILISATEUR	TEMPS DE LECTURE PAR RUBRIQUE	NOMBRE MOYEN DE PAGES PAR DOCUMENT	NOMBRE DE PAGES DANS LA BASE DE DONNEES		
			5000	10 000	15 000
EXHAUSTIVE	.25 sec	2	14	17	20
		5	10	13	15
	.50	2	13	17	20
		5	10	13	15
	1.00	2	7	9	9
		5	10	13	14
	2.00	2	5	8	9
		5	6	7	8
AUTOCESSANTE	.25 sec	2	14	17	20
		5	32	40	40
	.50	2	14	17	20
		5	10	13	15
	1.00	2	7	17	20
		5	10	13	15
	2.00	2	7	9	9
		5	6	13	14

[a] On suppose un débit de 16 multiplets par ligne du balayage télévisuel pour un débit réel de transmission des données de 241 920 multiplets/s.

TABLEAU 4

NOMBRE OPTIMAL DE RUBRIQUES PAR PAGE D'INDEX
DU TÉLIDON EN MODE DE RADIODIFFUSION AVEC TRANSMISSION
SUR LA TOTALITÉ D'UN CANAL[a] A 362 880 MULTIPLETS PAR SECONDE
(ANALYSE C)

MÉTHODE DE RECHERCHE DE L'UTILISATEUR	TEMPS DE LECTURE PAR RUBRIQUE	NOMBRE MOYEN DE PAGES PAR DOCUMENT	NOMBRE DE PAGES DANS LA BASE DE DONNÉES		
			5000	10 000	15 000
EXHAUSTIVE	.25 sec	2	14	17	20
		5	10	13	15
	.50	2	7	17	20
		5	10	13	15
	1.00	2	7	8	9
		5	6	7	14
	2.00	2	5	6	6
		5	4	7	7
AUTOCESSANTE	.25 sec	2	14	17	20
		5	32	40	40
	.50	2	14	17	20
		5	10	13	15
	1.00	2	7	17	20
		5	10	13	15
	2.00	2	7	8	9
		5	6	7	8

[a] On suppose un débit de 24 multiplets par ligne du balayage télévisuel pour un débit réel de transmission des données de 362 880 multiplets/s.

TABLEAU 5

TEMPS DE RECHERCHE PREVUS POUR LE
SYSTEME TELIDON EN MODE DE RADIODIFFUSION
AVEC TRANSMISSION SUR LA TOTALITE D'UN CANAL

DEBIT REEL DE TRANSMISSION DES DONNEES[a]	DIMENSION DE LA BASE DE DONNEES[b]	TEMPS DE RECHERCHE PREVUS[c]	TEMPS DE RECHERCHE PREVUS ($a \leq 10$)[d]
241,920	5,000	30 - 91	33 - 91
	10,000	54 - 135	65 - 135
	15,000	81 - 173	97 - 173
362,880	5,000	23 - 79	24 - 79
	10,000	41 - 113	47 - 113
	15,000	59 - 140	68 - 140

[a]Mesuré en multiplets/s. Le débit réel de transmission des données est la vitesse à laquelle l'information est diffusée sur un canal de télévision. Ce débit est égal au produit de a) BL, nombre de multiplets de données par ligne du balayage télévisuel, b) NL, nombre de lignes utiles du balayage dans une trame et c) C, vitesse du cycle en trames/s. Pour une transmission sur la totalité d'un canal, NL = 252 lignes utiles du balayage et C = 60 trames/s. Pour BL = 16, le débit réel de transmission des données = BL x NL x C = 241 920 multiplets/s. Pour BL = 24, le débit réel de transmission des données = 362 880 multiplets/s.

[b]Nombre total de pages de la base de données, y compris les pages d'index et les pages de documentation.

[c]Mesurés en secondes. Les temps de recherche indiqués représentent l'éventail des temps de recherche (ST) optimums.

[d]Mesurés en secondes. Les temps de recherche signalés représentent l'éventail des temps de recherche optimums (localement) lorsqu'il y a au plus 10 rubriques par page.

Figure 1: Nombre optimal de rubriques par page d'index en fonction de la dimension de la base de données du Télidon en mode de radiodiffusion.

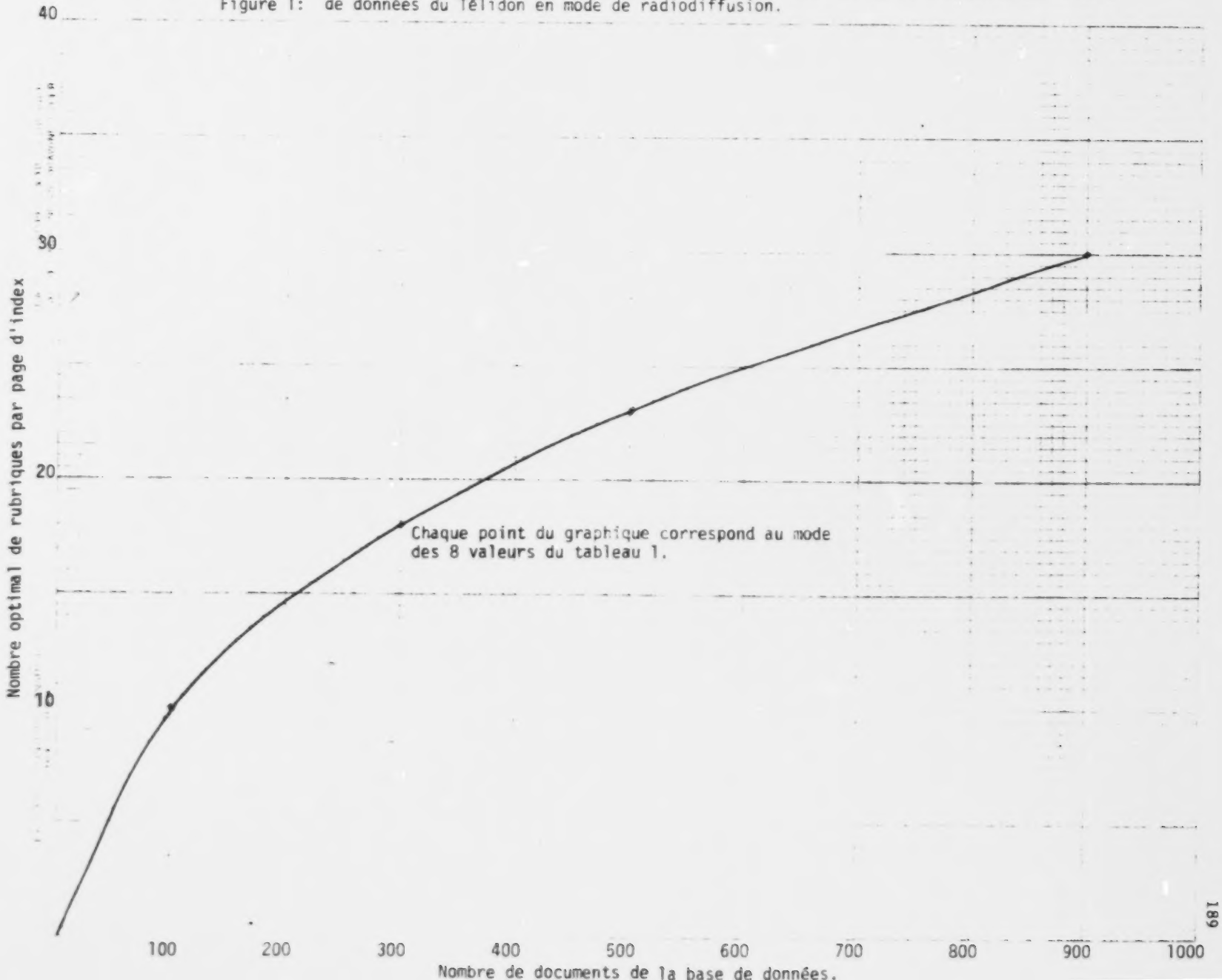


FIGURE 2a: Temps de recherche en fonction du nombre de rubriques par page d'index (pour 100 pages de documentation)

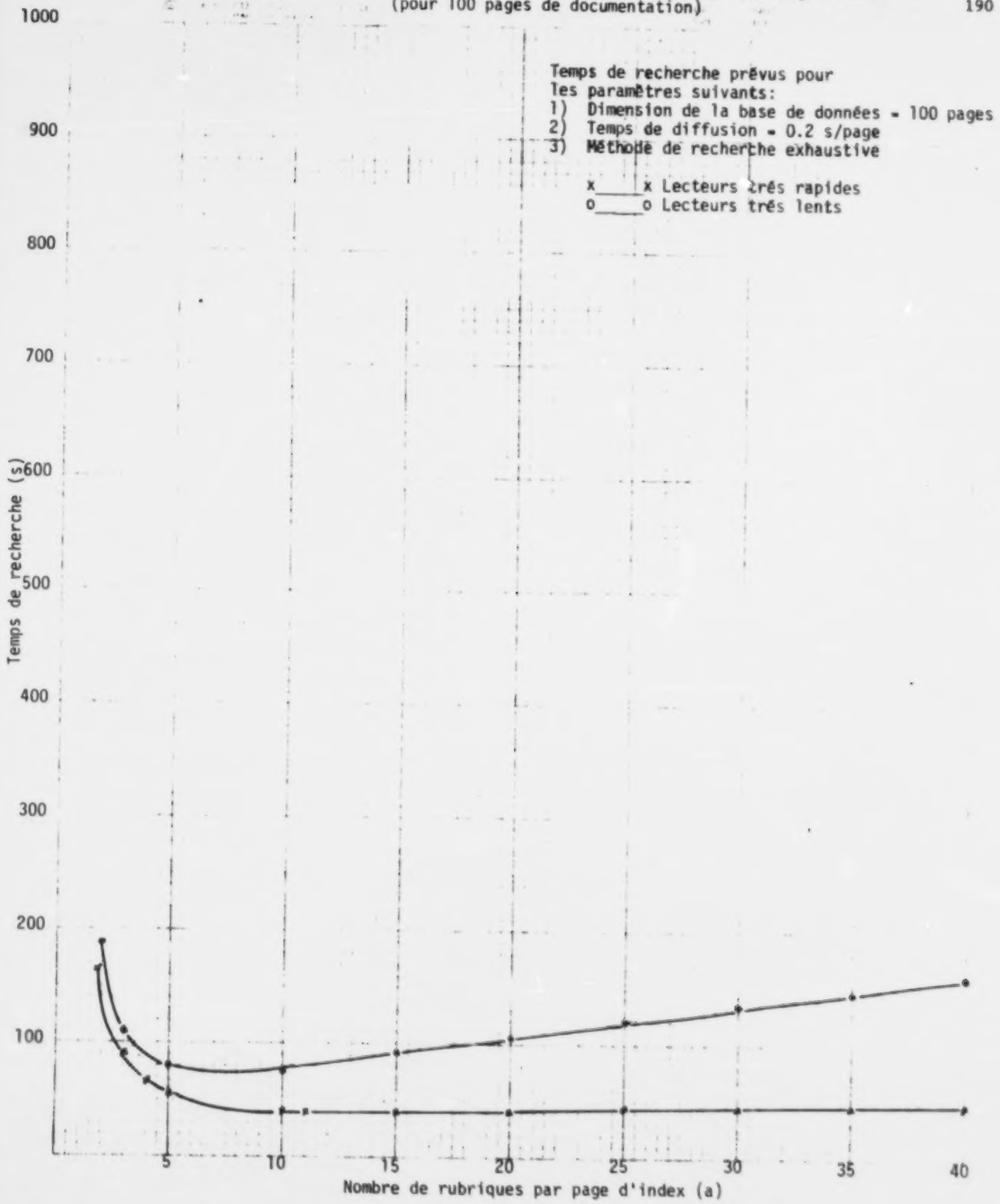


FIGURE 2b: Search Time As a Function of the Number of Alternatives Per Index Page
(for 300 document pages)

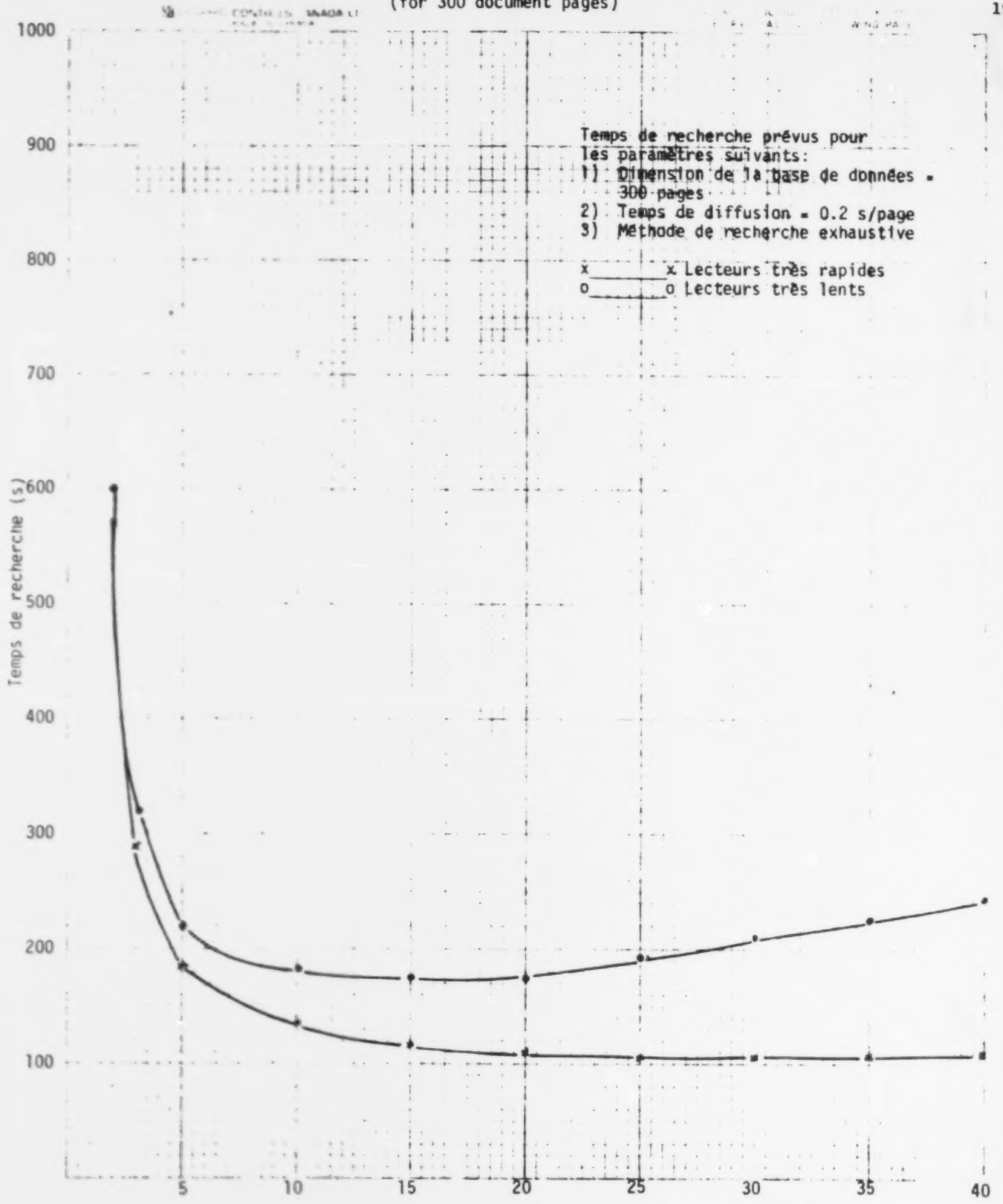


FIGURE 2c: Temps de recherche en fonction du nombre de rubriques par page d'index (pour 500 pages de documentation)

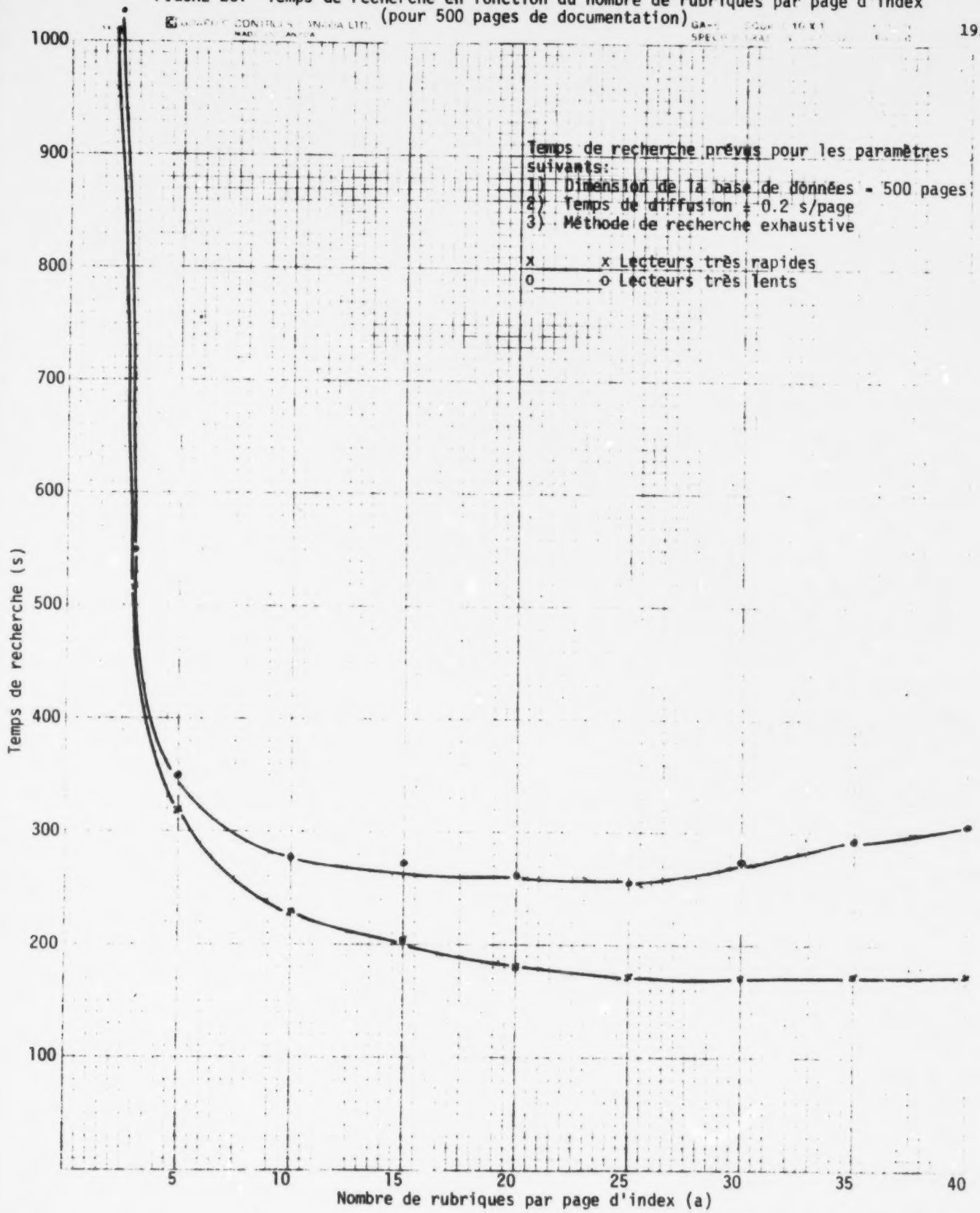


FIGURE 2d: Search Time As a Function of the Number of Alternatives Per Index Page (for 900 document pages)

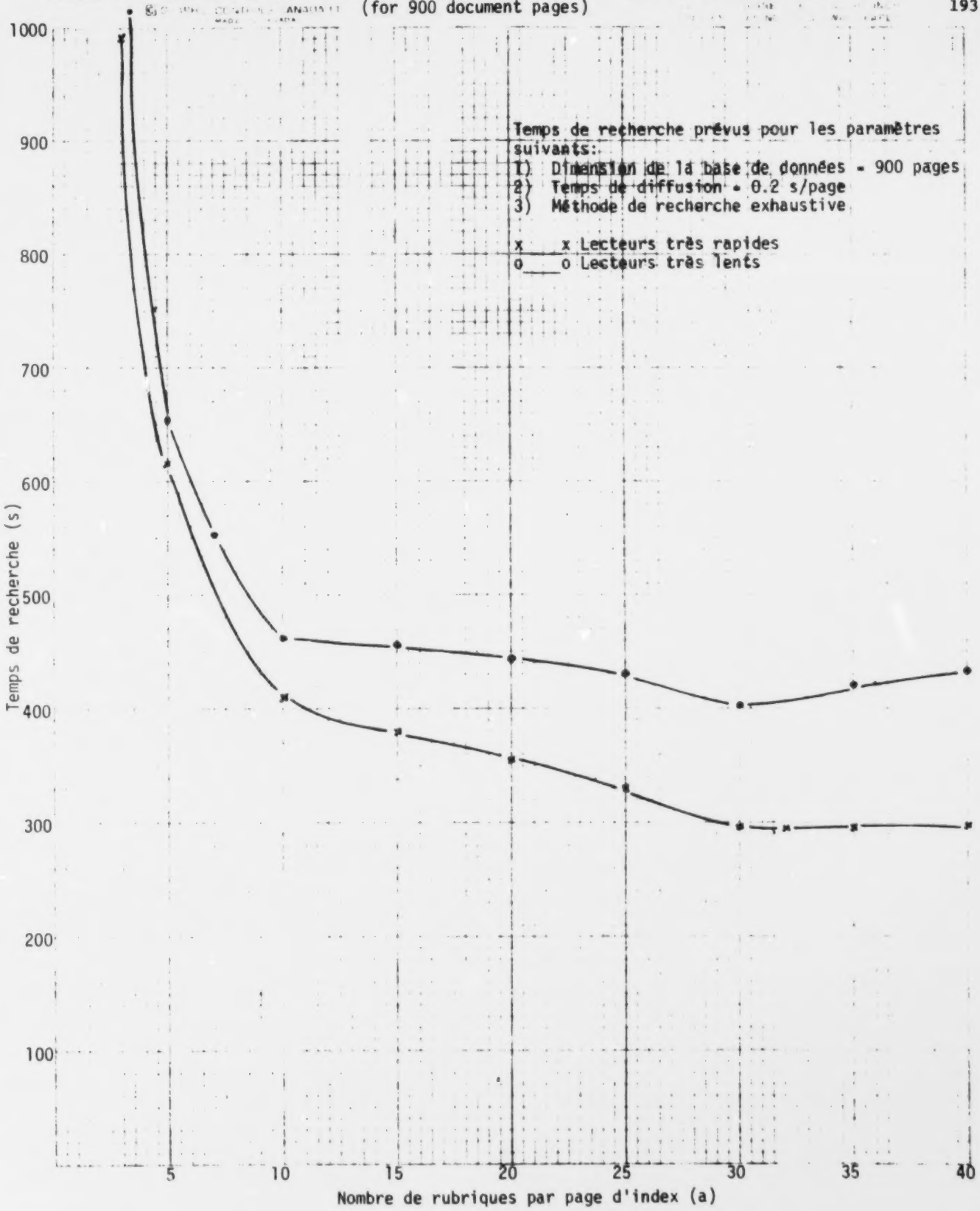


FIGURE 3a: Temps de recherche en fonction du nombre de rubriques par page d'index (pour 100 pages de documentation)

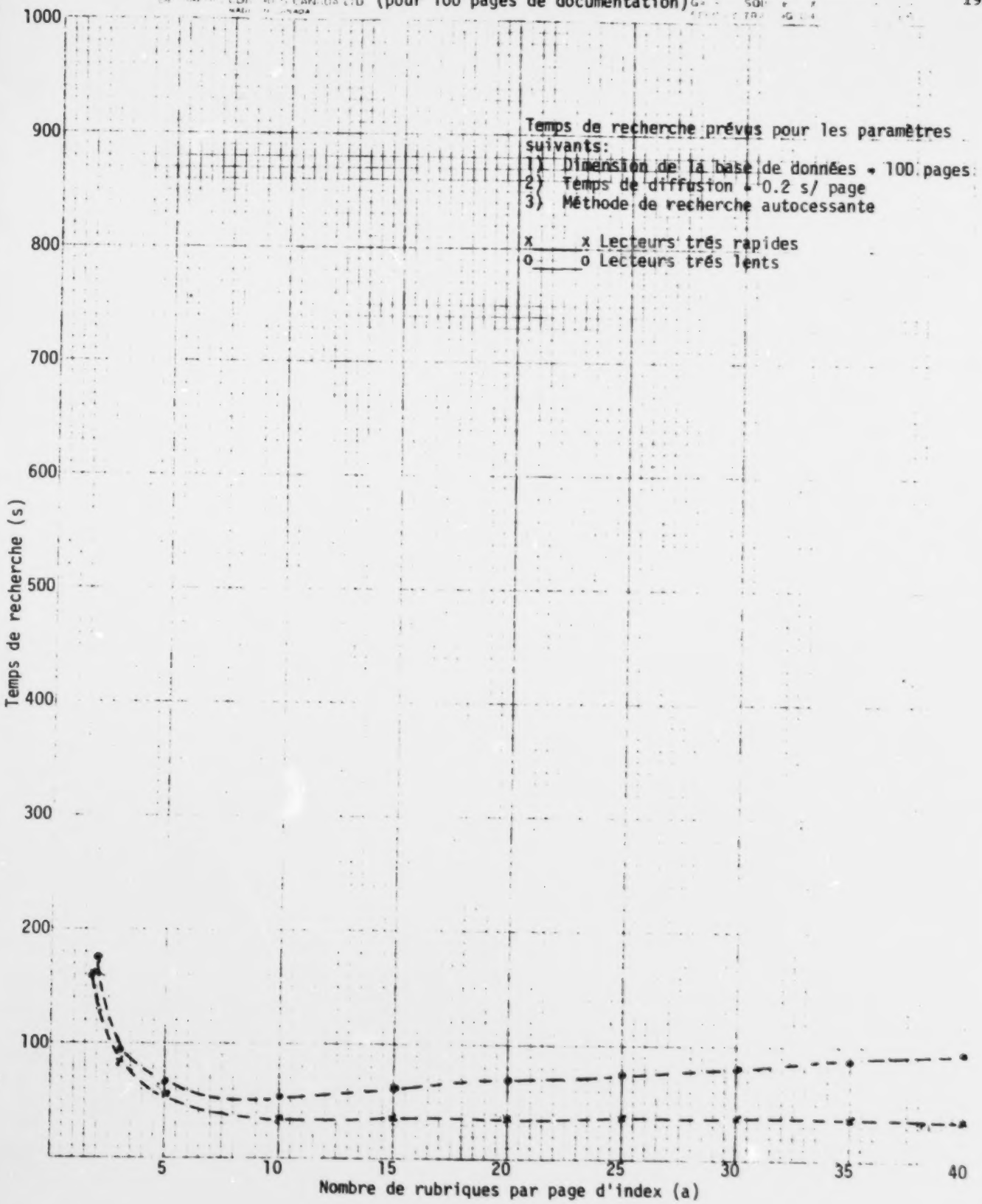


FIGURE 3b: Temps de recherche en fonction du nombre de rubriques par page d'index (pour 300 pages de documentation)

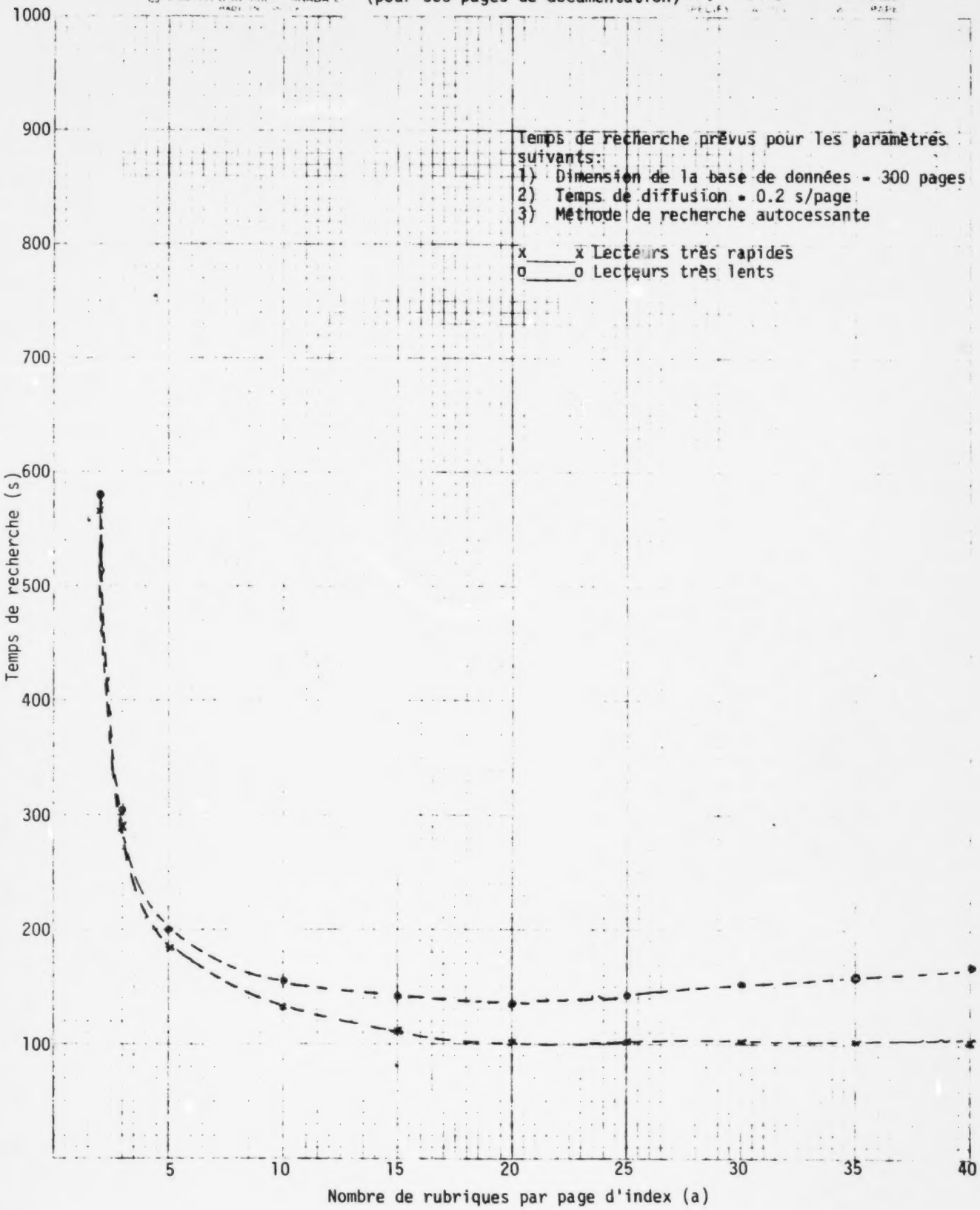


FIGURE 3c: Temps de recherche en fonction du nombre de rubriques par page d'index (pour 500 pages de documentation)

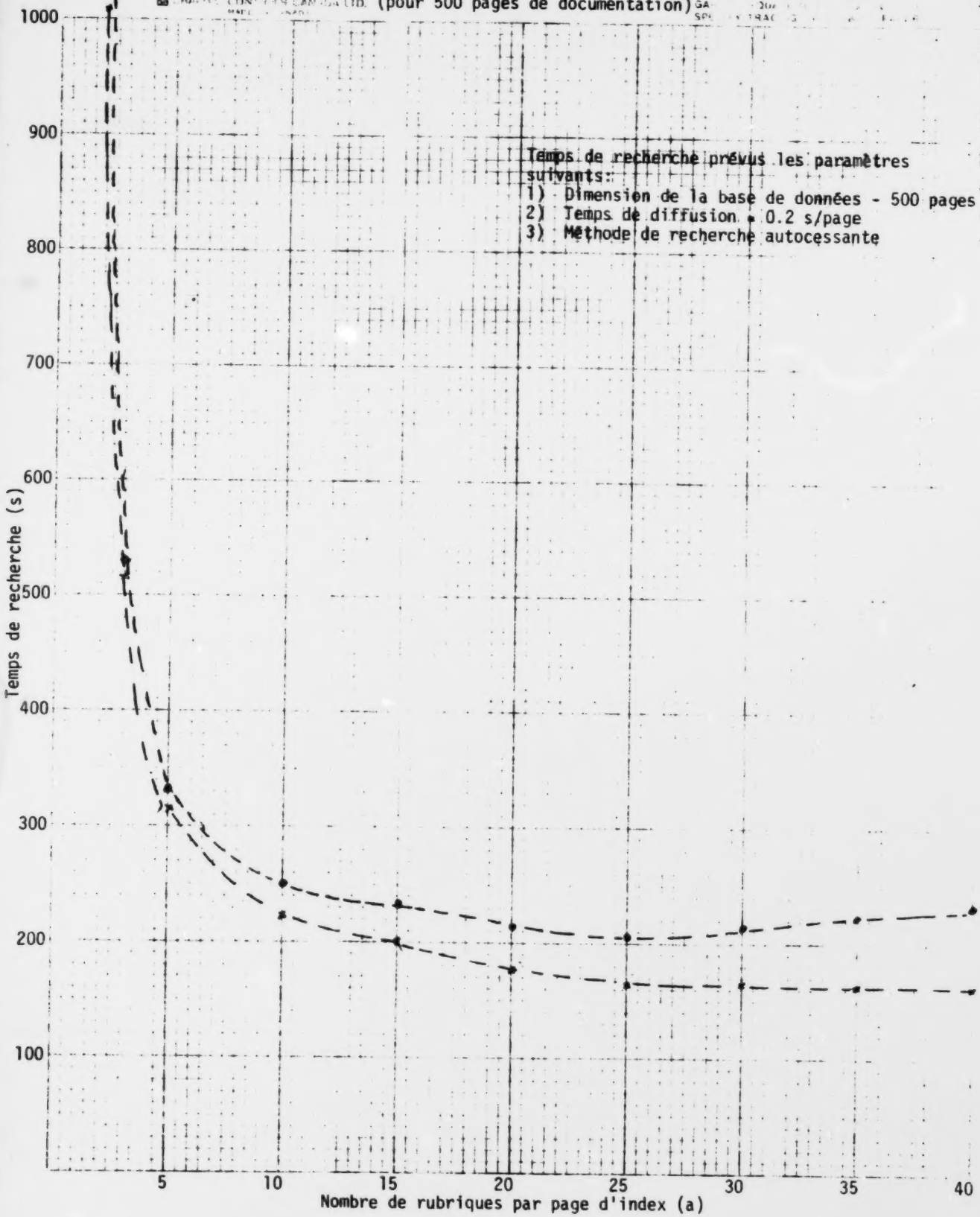


FIGURE 3d: Temps de recherche en fonction du nombre de rubriques par page d'index (pour 900 pages de documentation)

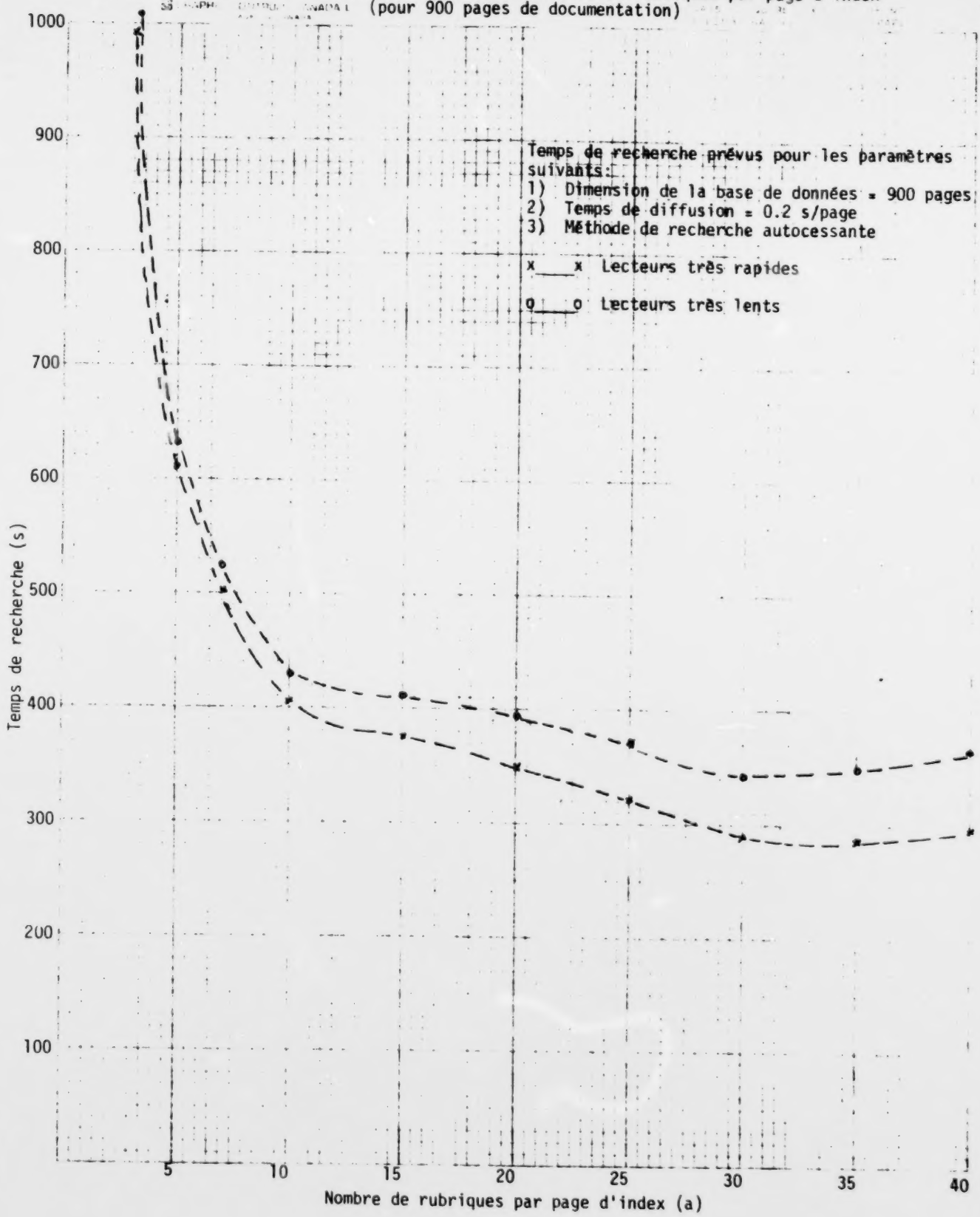


FIGURE 4a: Exemple d'une page d'index du Télidon offrant un choix de 20 rubriques

1. Transport
2. Avion
3. Bateau
4. Autobus
5. Traversier
6. OC Transpo
7. Taxi
8. Train
9. Services spéciaux
10. Index 1
11. Alimentation
12. Pour les bébés
13. Finance
14. Assurances
15. Index logement
16. Auto-assistance
17. Équipement automoteur
18. Outils
19. Locations
20. Appartements

FIGURE 4b: Exemple d'une page du Télidon offrant un choix de 40 rubriques

1. Appartements
2. Duplex
3. Condominiums
4. Maisons en rangée
5. Maisons
6. Sous-locations
7. Chambre et pension
8. Rue Rideau
9. Secteur du marché
10. Basse-ville
11. Ottawa-Ouest
12. Ottawa-Est
13. Ottawa-Sud
14. Banlieue
15. Beaux-Arts
16. Danse
17. Design
18. Musique
19. Photographie
20. Spectacles
21. Dîner en ville
22. Clubs
23. Médias
24. Arts
25. Musées
26. Index 1
27. Index 2
28. Services spéciaux
29. District
30. De la musique
31. Des clubs
32. Budger alimentation
33. Nutrititon
34. Liste d'épicerie
35. Mesures métriques
36. Produits sans marque
37. Budget alimentation
38. Budget produits laitiers
39. Budget fruits
40. Budget viande

FIGURE 4c: Exemple d'une page du Télidon offrant un choix de 60 rubriques

1. Ottawa-Est
2. Glebe
3. Rockcliffe
4. Ottawa-Sud
5. Vanier
6. Ottawa-Ouest
7. Investissements
8. Carrières
9. Economie
10. National
11. Local
12. District
13. Plats-minute
14. 24 heures par jour
15. Prévisions
16. Pièces d'automobile
17. Isolation
18. Température
19. Carte
20. Locations
21. Propriété
22. Animaux
23. Services
24. Mexicain
25. Naturel
26. Extérieur
27. Chaussures
28. Crédit
29. Couvertures
30. Serviettes
31. Linge
32. Oreillers
33. Maisons
34. Alberta
35. C.-B.
36. Manitoba
37. Ontario
38. Québec
39. Î.-P.-É.
40. D'après la marque
41. A l'année
42. Télévision
43. Piano
44. Orgue
45. Articles
46. Antiquités
47. Équipement
48. Mobilier
49. Bois de chauffage
50. Perspective
51. Magasin d'alimentation
52. Écoles
53. Collèges
54. Bibliothèques
55. Voitures de sport
56. Kystes
57. Allergies
58. Maux de tête
59. Insomnie
60. Psoriasis

FIGURE 4d: Exemple d'une page du Télidon offrant un choix de 80 rubriques

1.	Ménage	21.	Kystes	41.	H-K	61.	Radio
2.	Sénat	22.	Télévision	42.	L-P	62.	Stéréo
3.	Cote	23.	Piano	43.	V-Z	63.	Fer
4.	Laveuse	24.	Orgue	44.	Q-T	64.	Elmdale
5.	Incendie	25.	Camions	45.	Bifteck	65.	Sécheuse
6.	Police	26.	C.-B.	46.	Suisse	66.	Batteur
7.	Clubs	27.	Québec	47.	Folk	67.	Médicaments
8.	Médias	28.	Î.-P.-É.	48.	Jazz	68.	Soupe
9.	Arts	29.	Maisons	49.	Rock	69.	Perruques
10.	Artisanat	30.	Linge	50.	Bar	70.	Asthme
11.	Danse	31.	Serviettes	51.	Café	71.	Ulcères
12.	Design	32.	Vanier	52.	Pub	72.	Cancer
13.	Alimentation	33.	Glebe	53.	Comédie	73.	Patinage
14.	Outils	34.	Crédit	54.	Drame	74.	Ski
15.	Avion	35.	Chaussures	55.	Épouvante	75.	Romans
16.	Bateau	36.	Local	56.	Cote	76.	Porcelaine
17.	Autobus	37.	Température	57.	Mines	77.	Timbres
18.	Traversier	38.	Carte	58.	Animaux de compagnie	78.	Pièces de monnaie
19.	Taxi	39.	A-C	59.	Chevaux	79.	Autos
20.	Train	40.	D-G	60.	Livres	80.	Volaille

FIGURE 5a: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

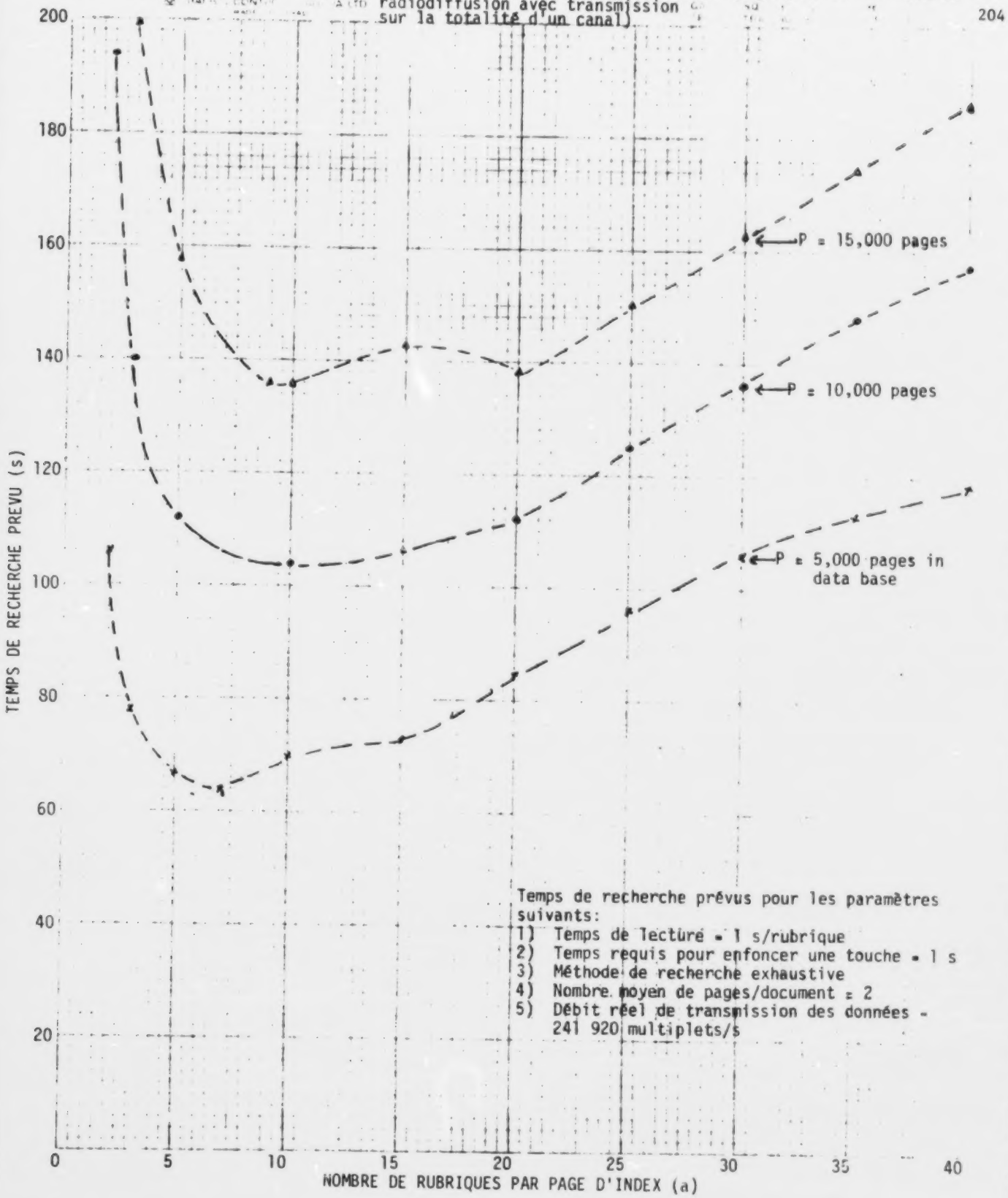


FIGURE 5b: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

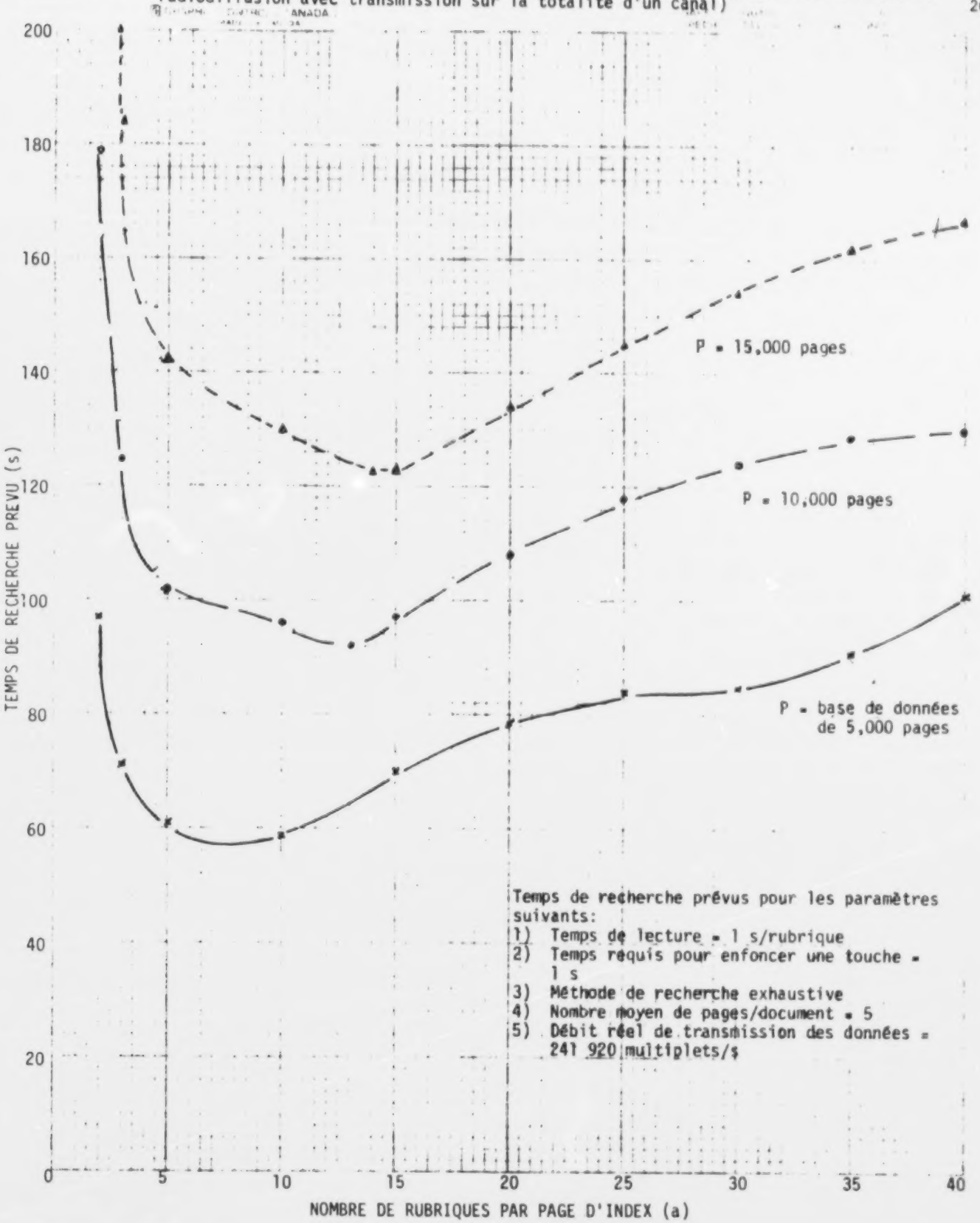


FIGURE 5c: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

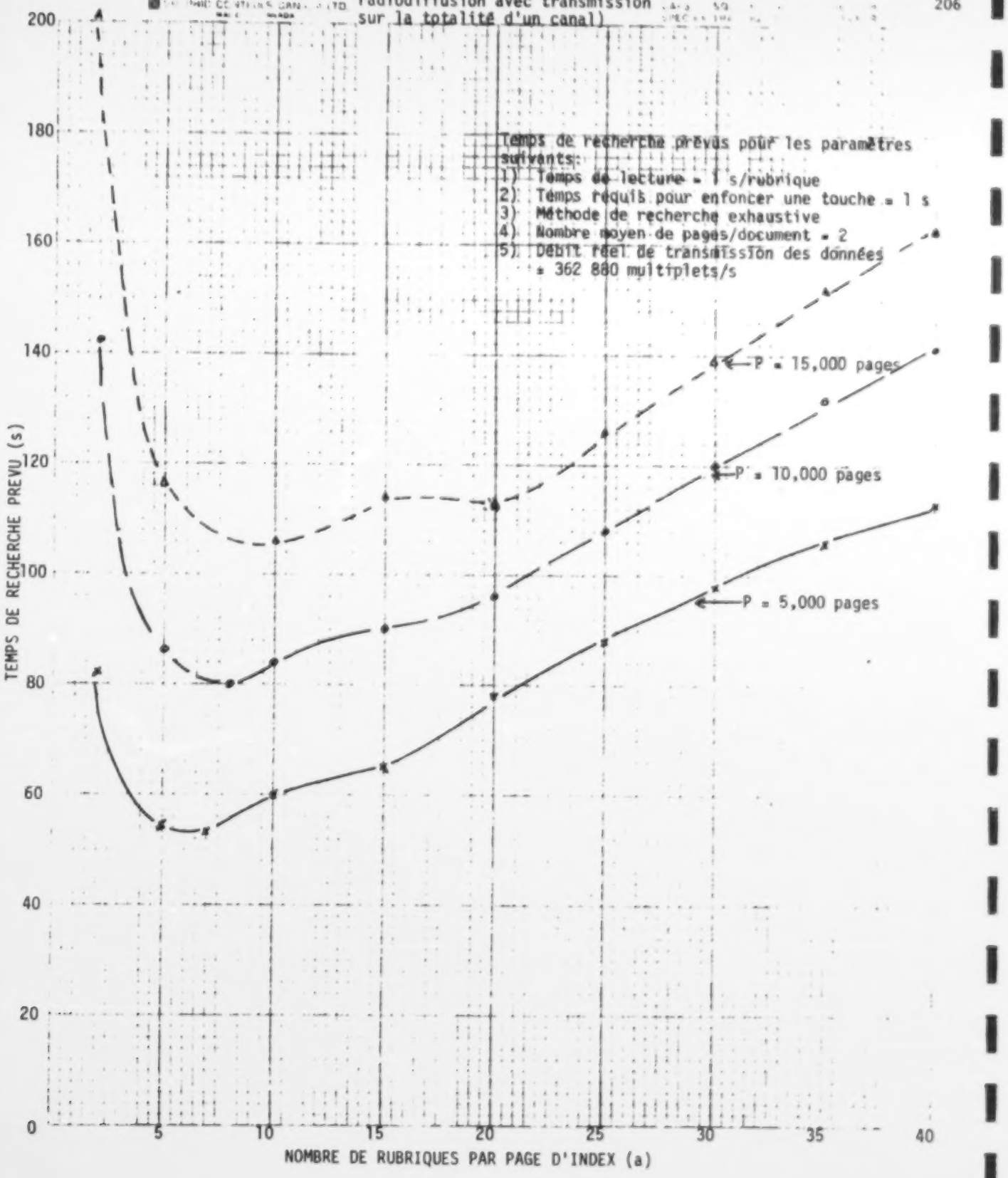


FIGURE 5d: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

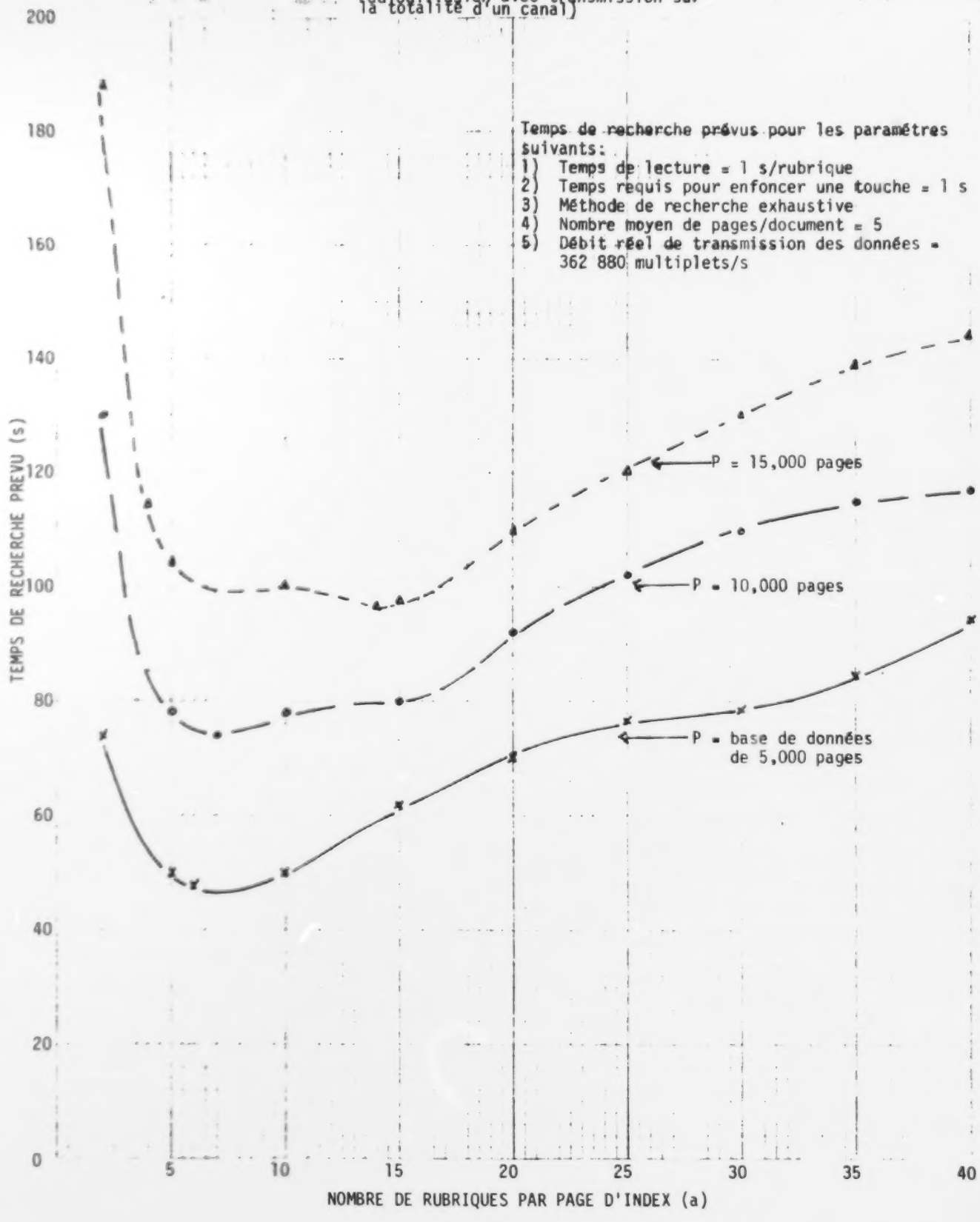


FIGURE 5e: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

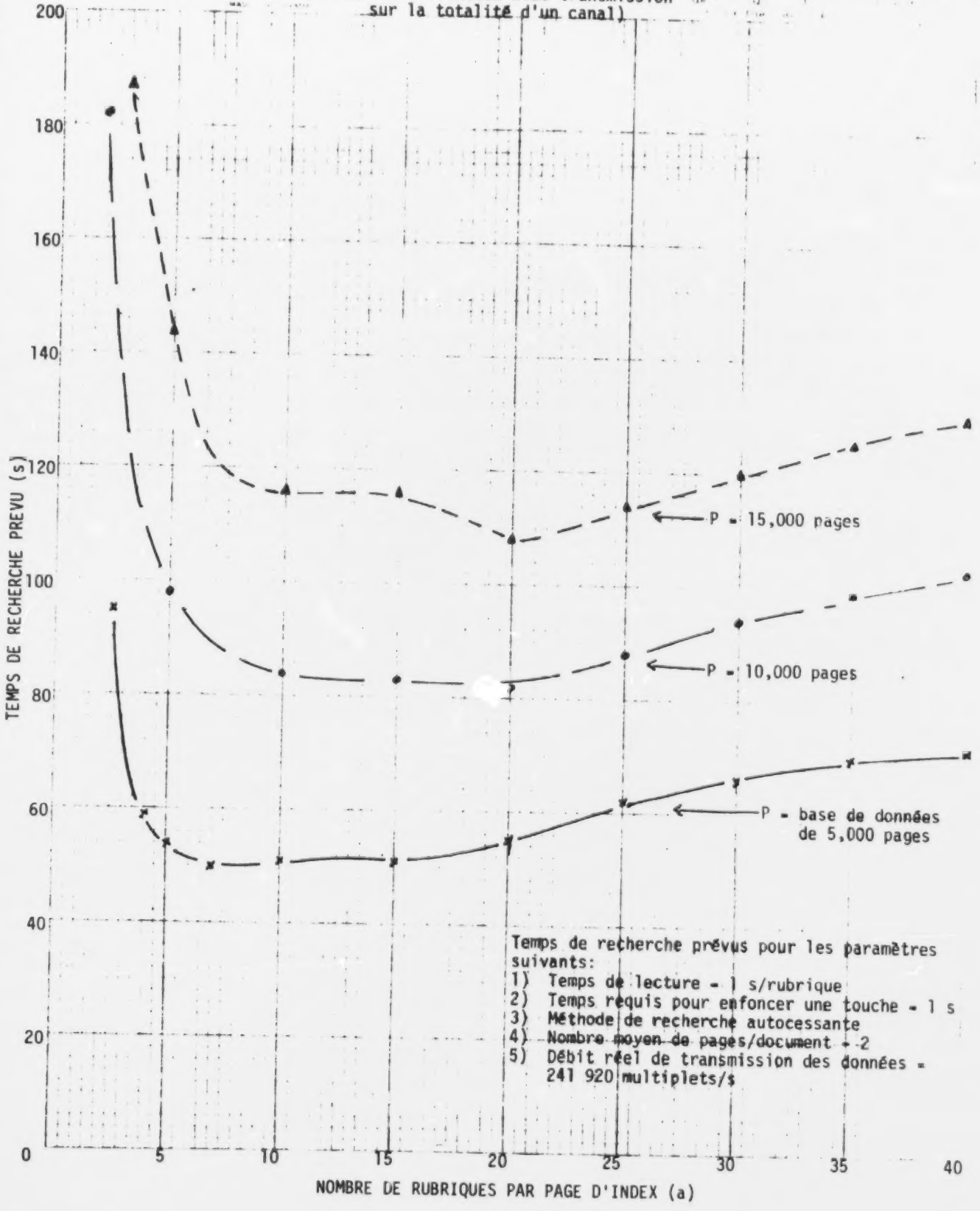


FIGURE 5f: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

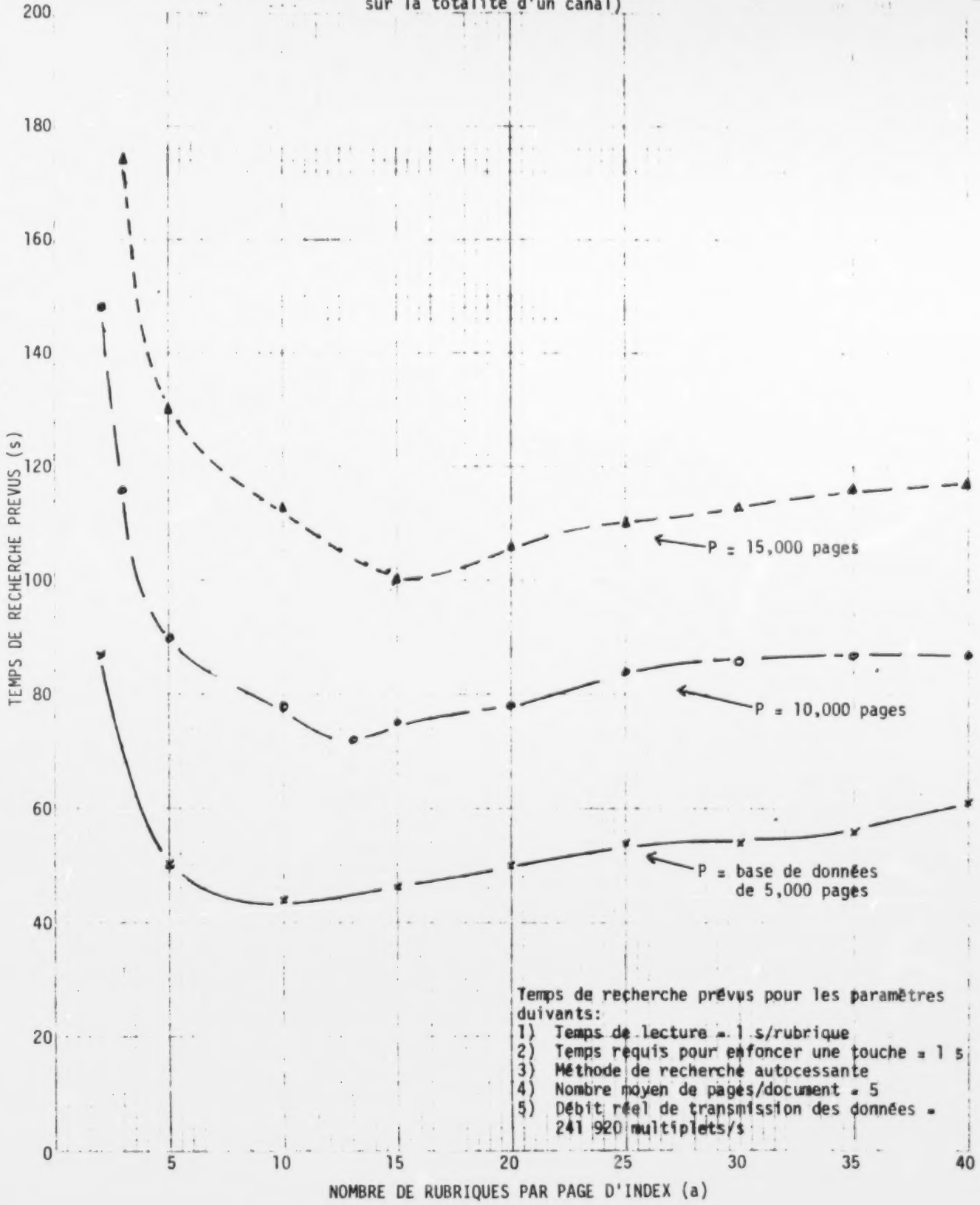


FIGURE 5g: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

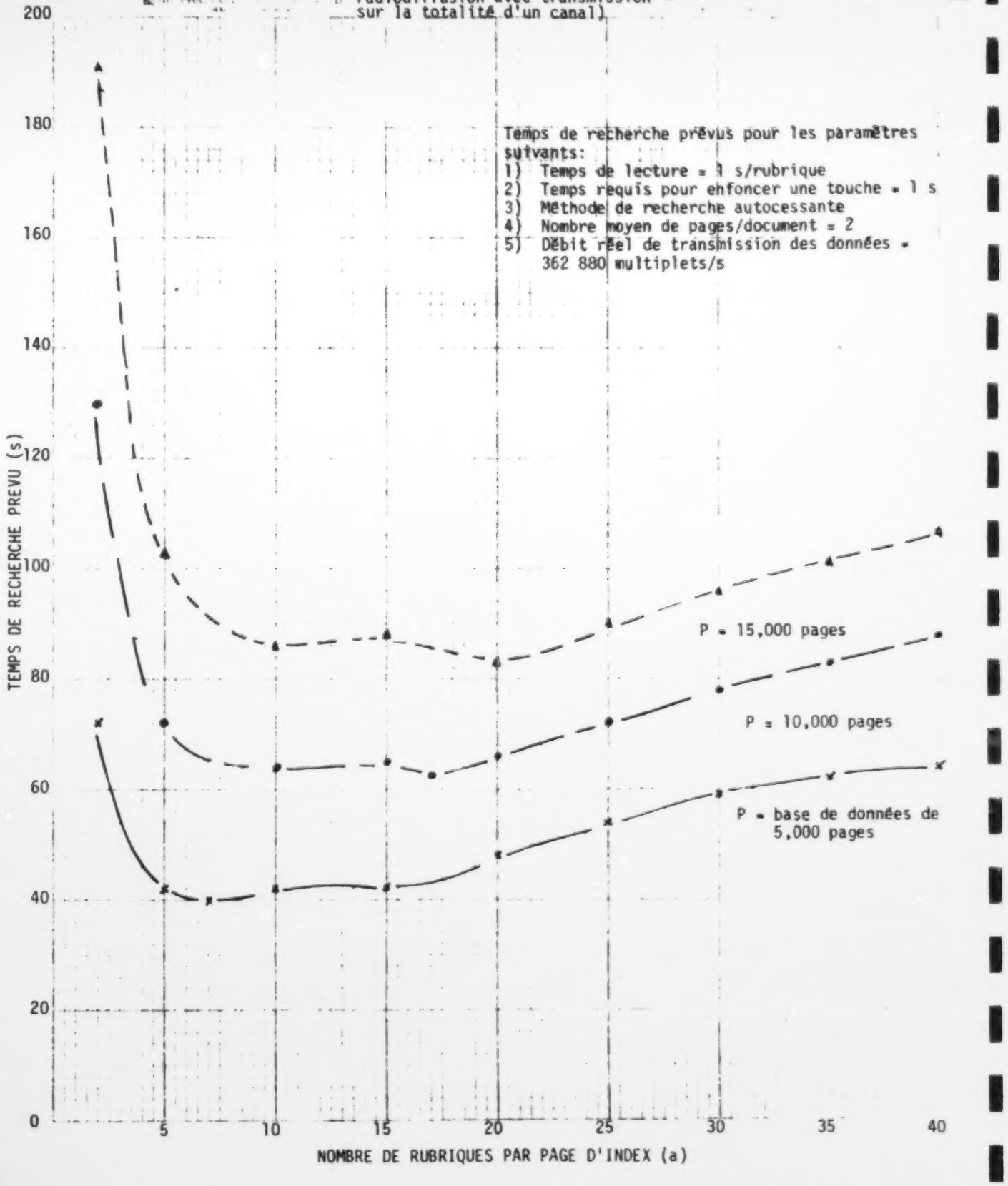


FIGURE 5h: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

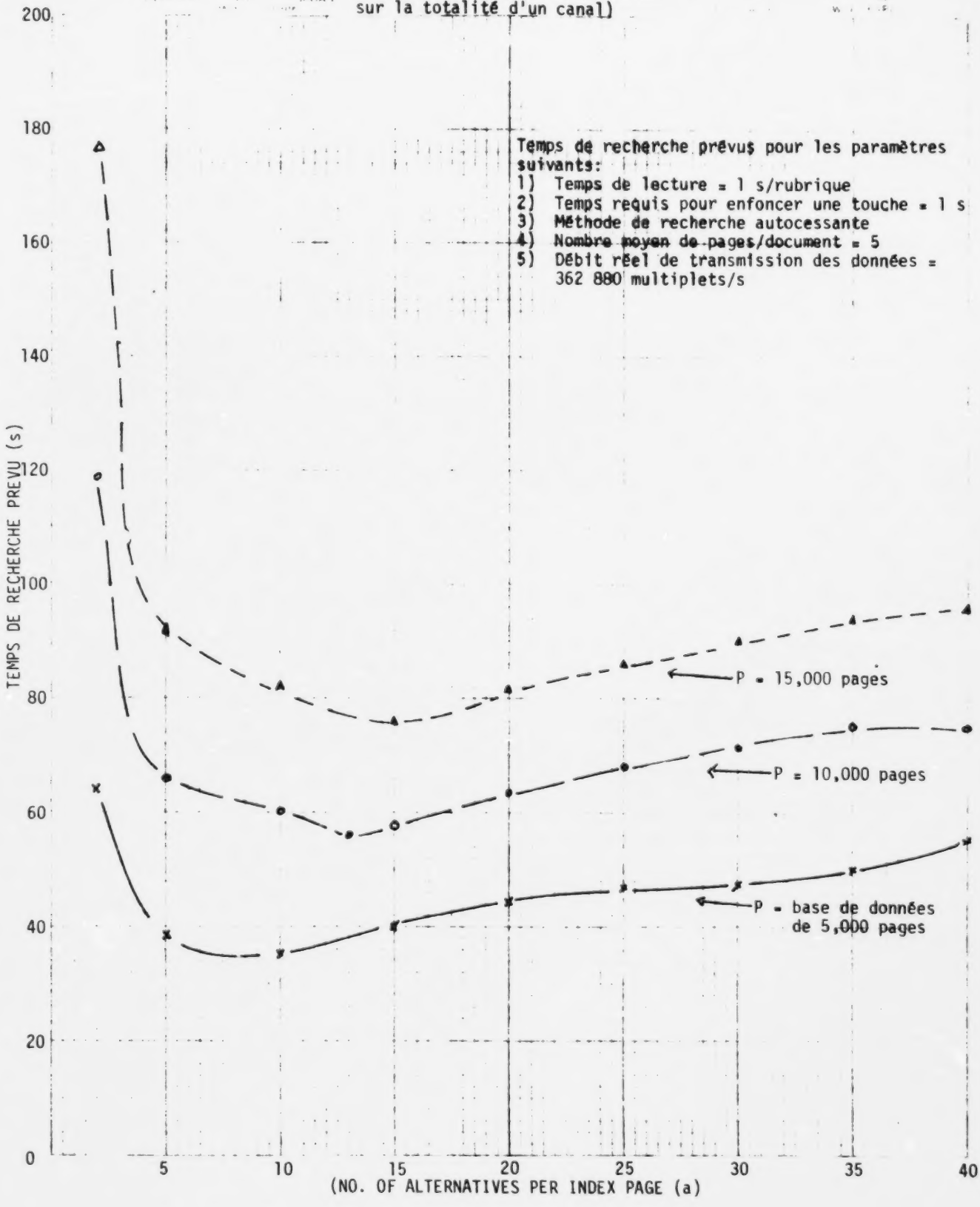


FIGURE 6: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

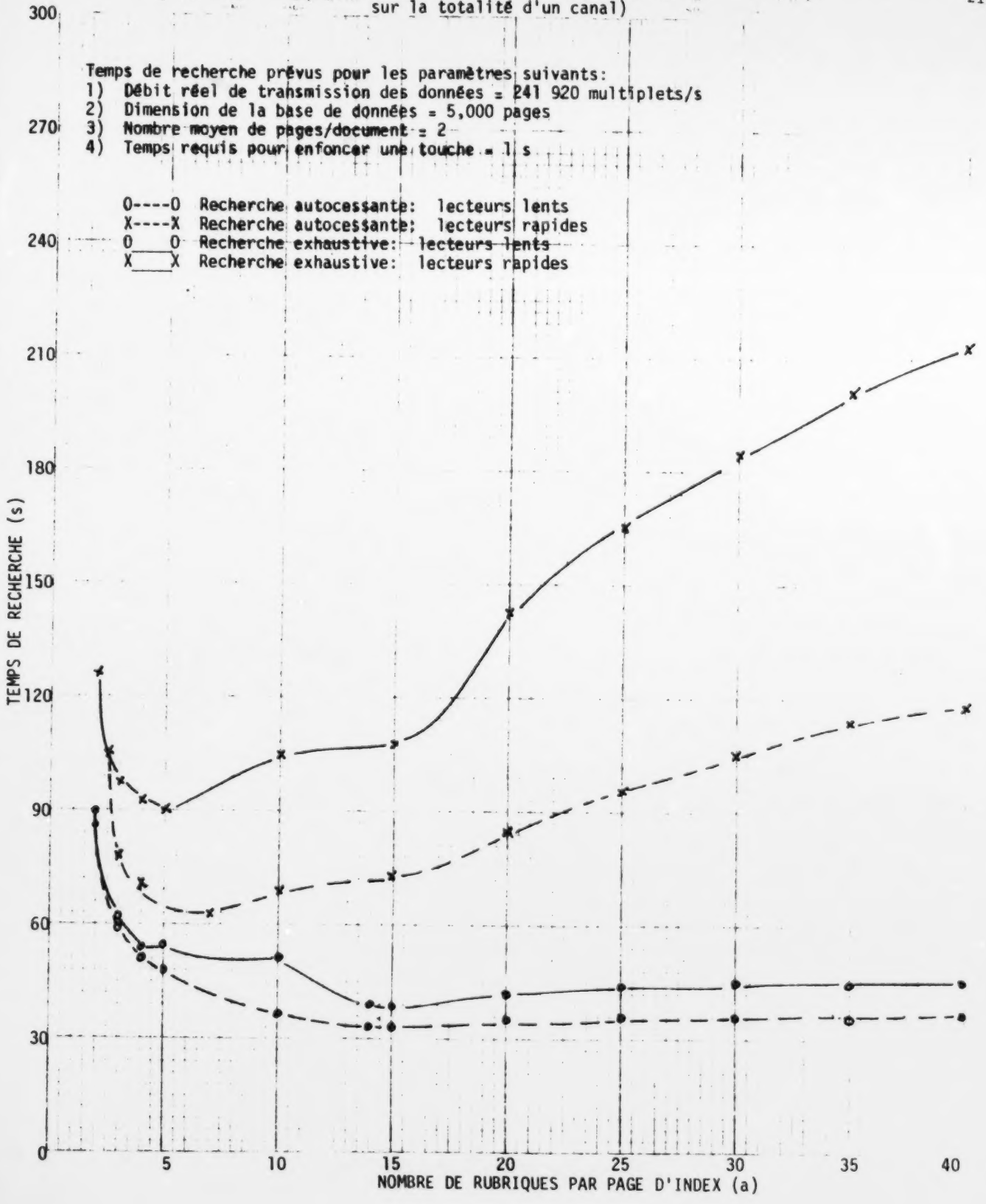


FIGURE 7: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

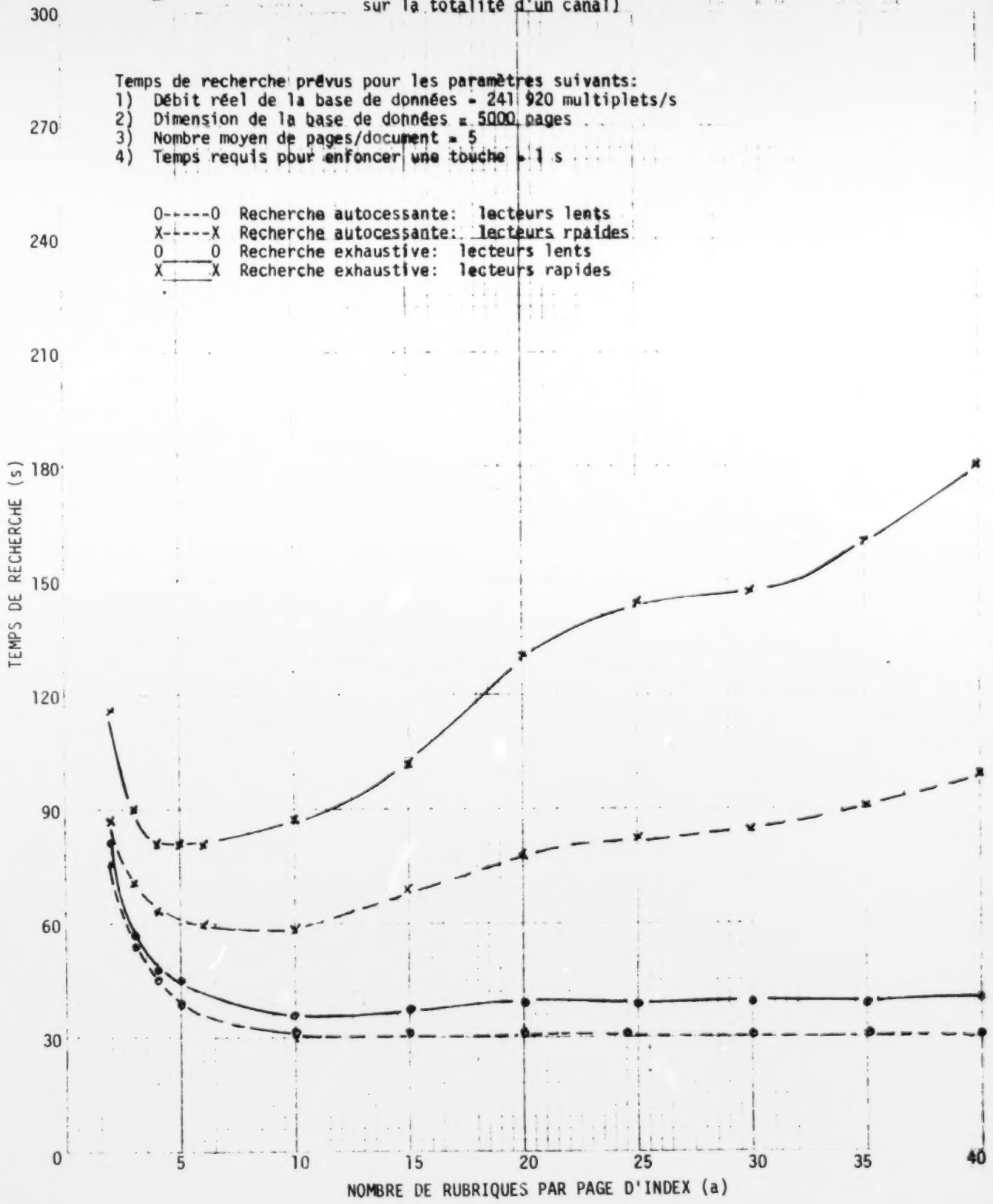


FIGURE 8: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

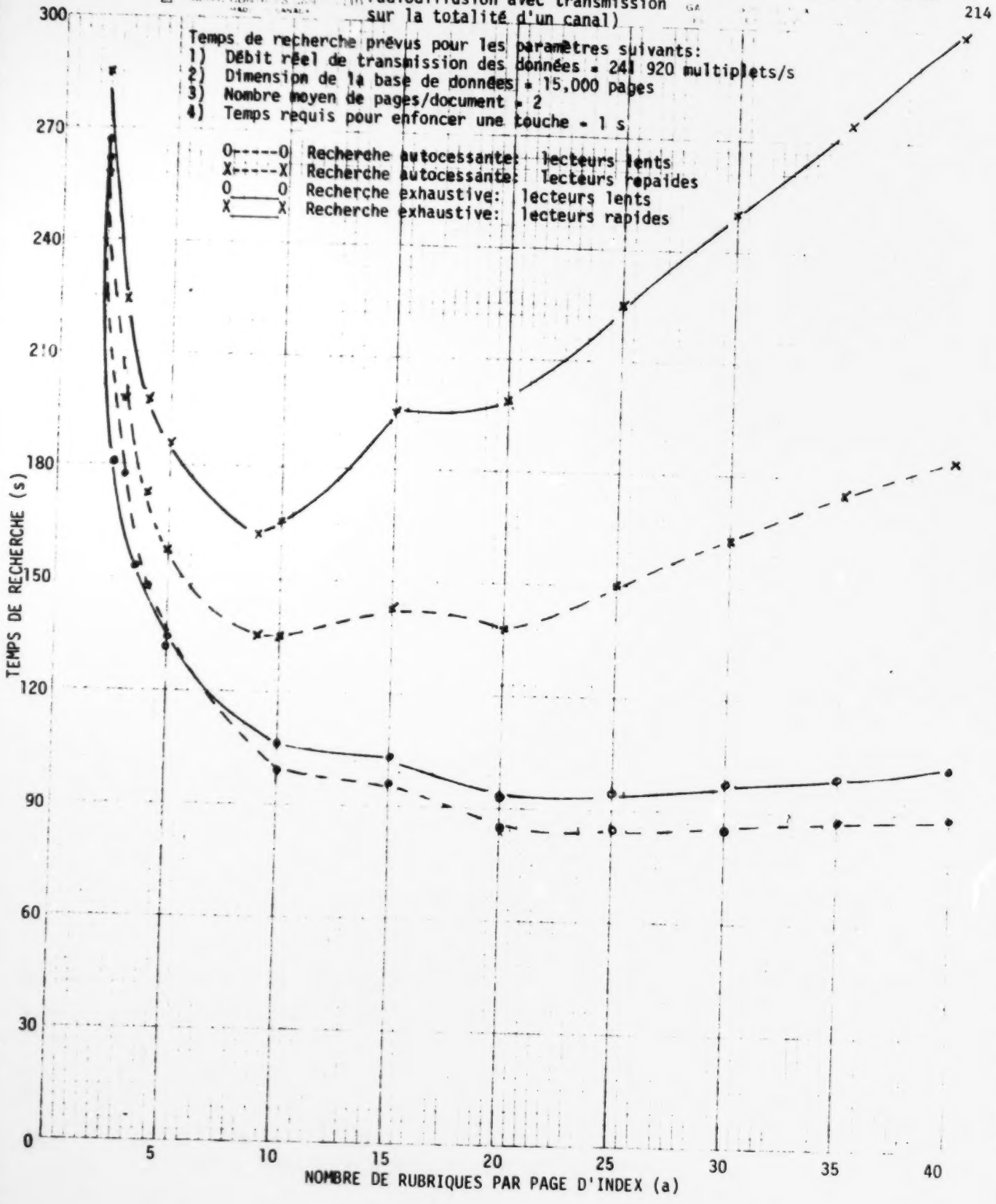


FIGURE 9: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

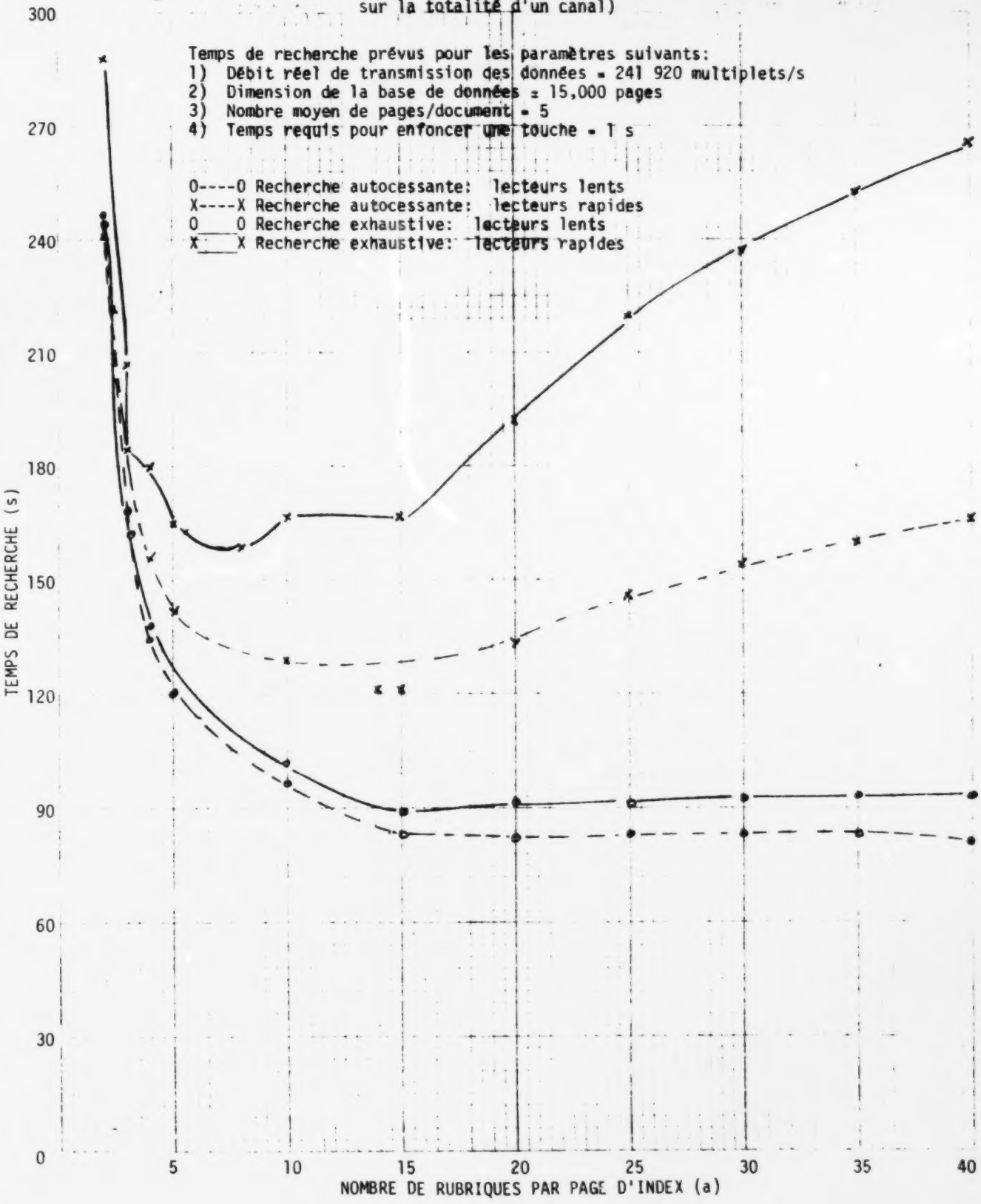


FIGURE 10: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

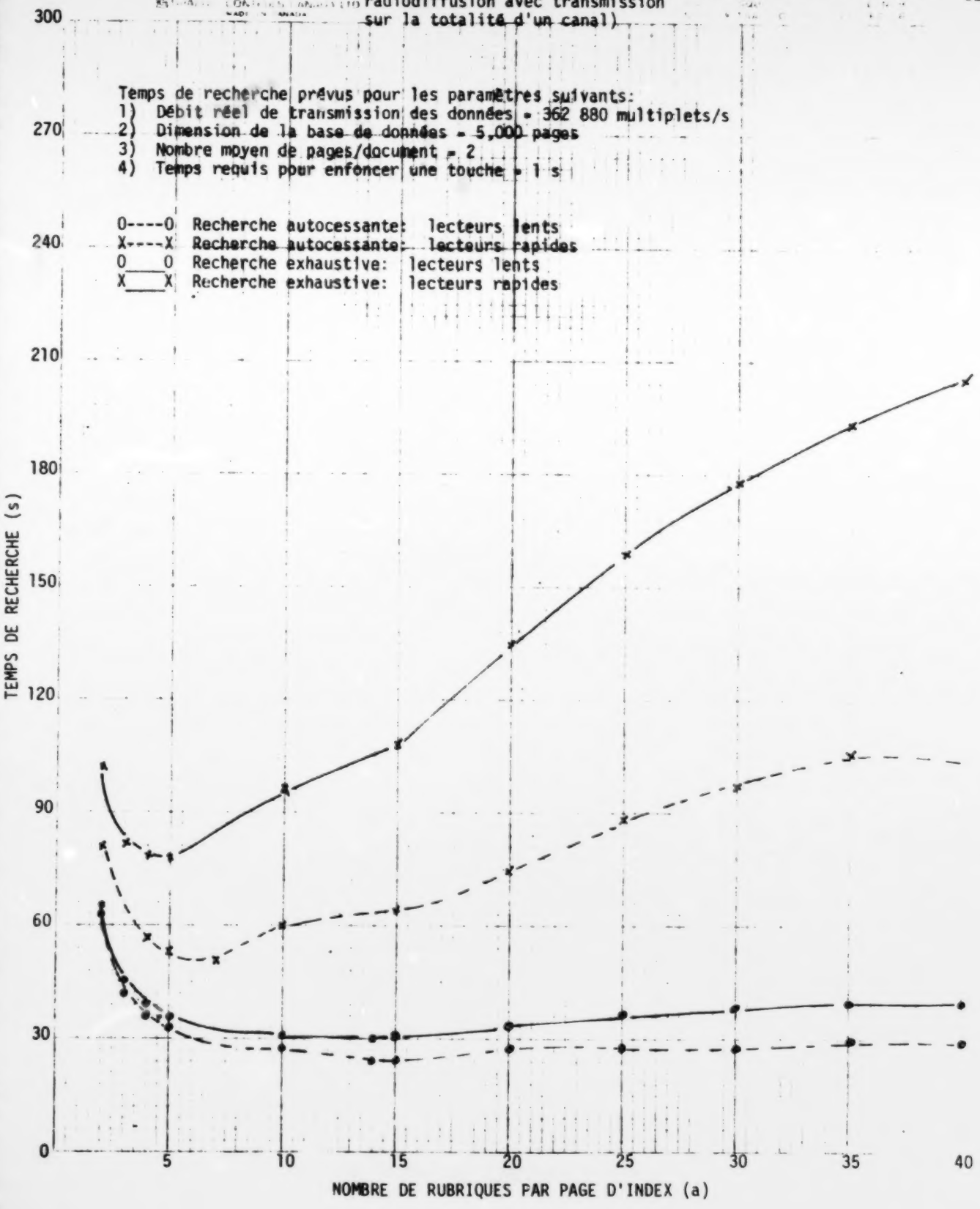


FIGURE 11: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

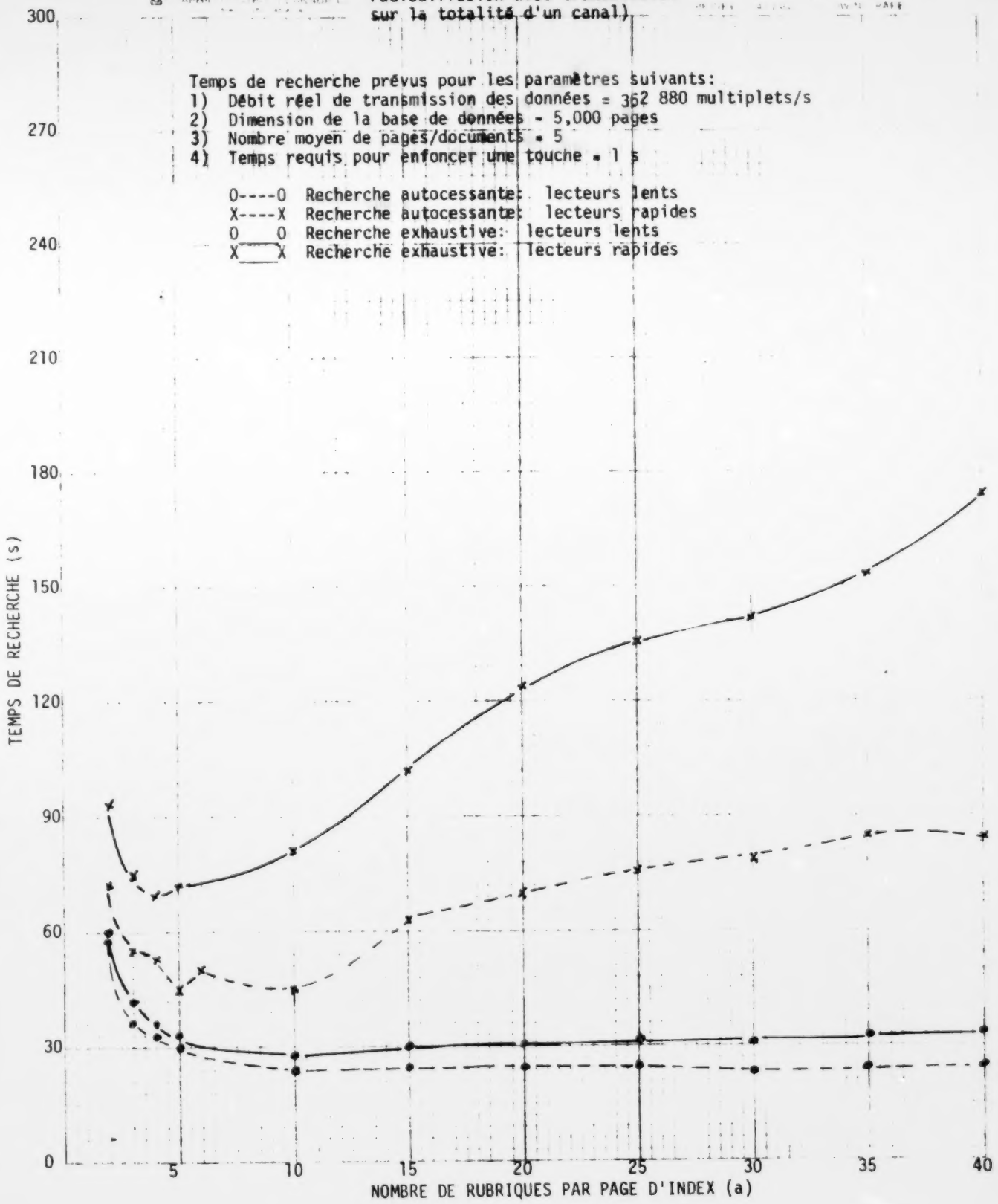


FIGURE 12: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

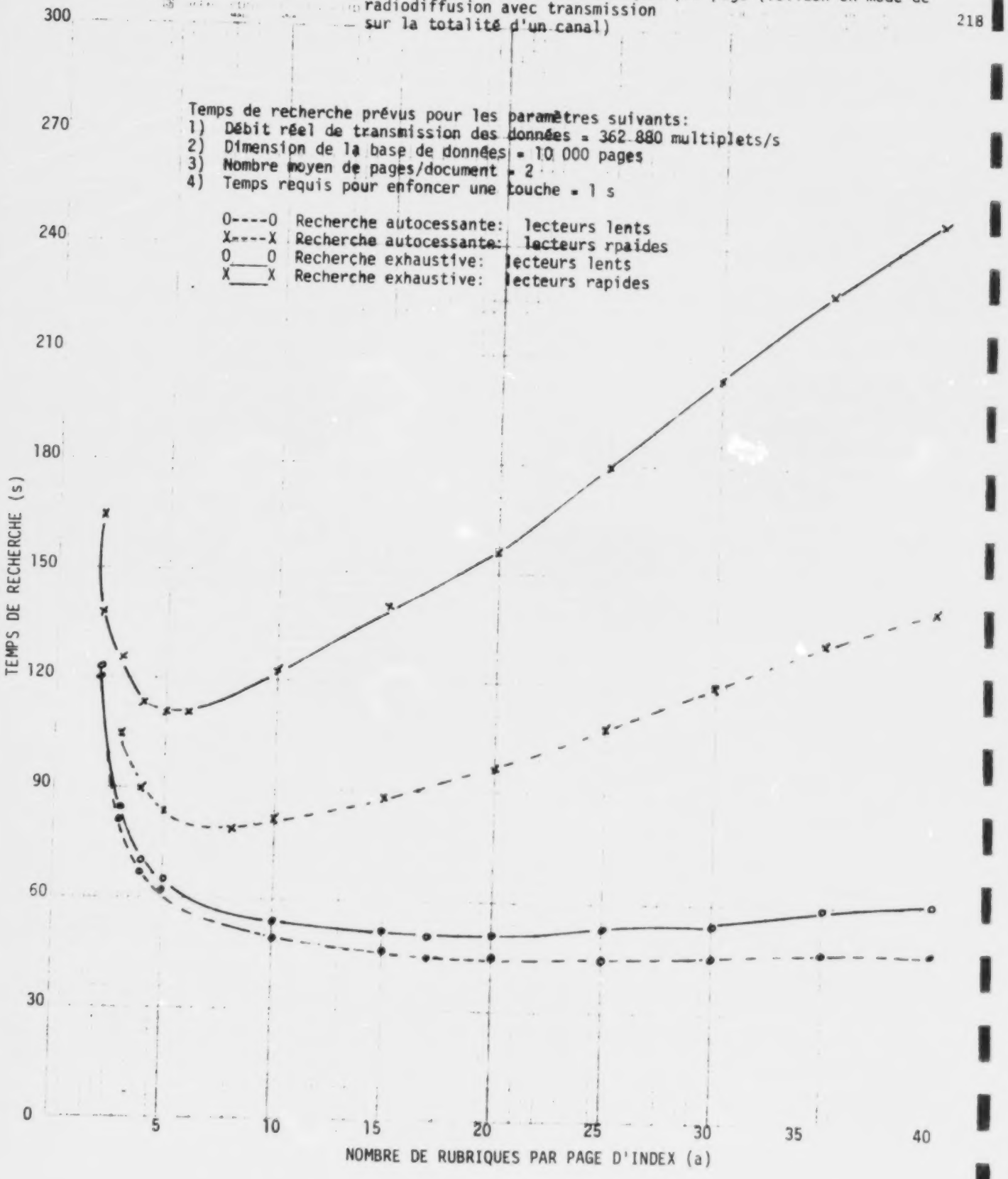


FIGURE 13: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

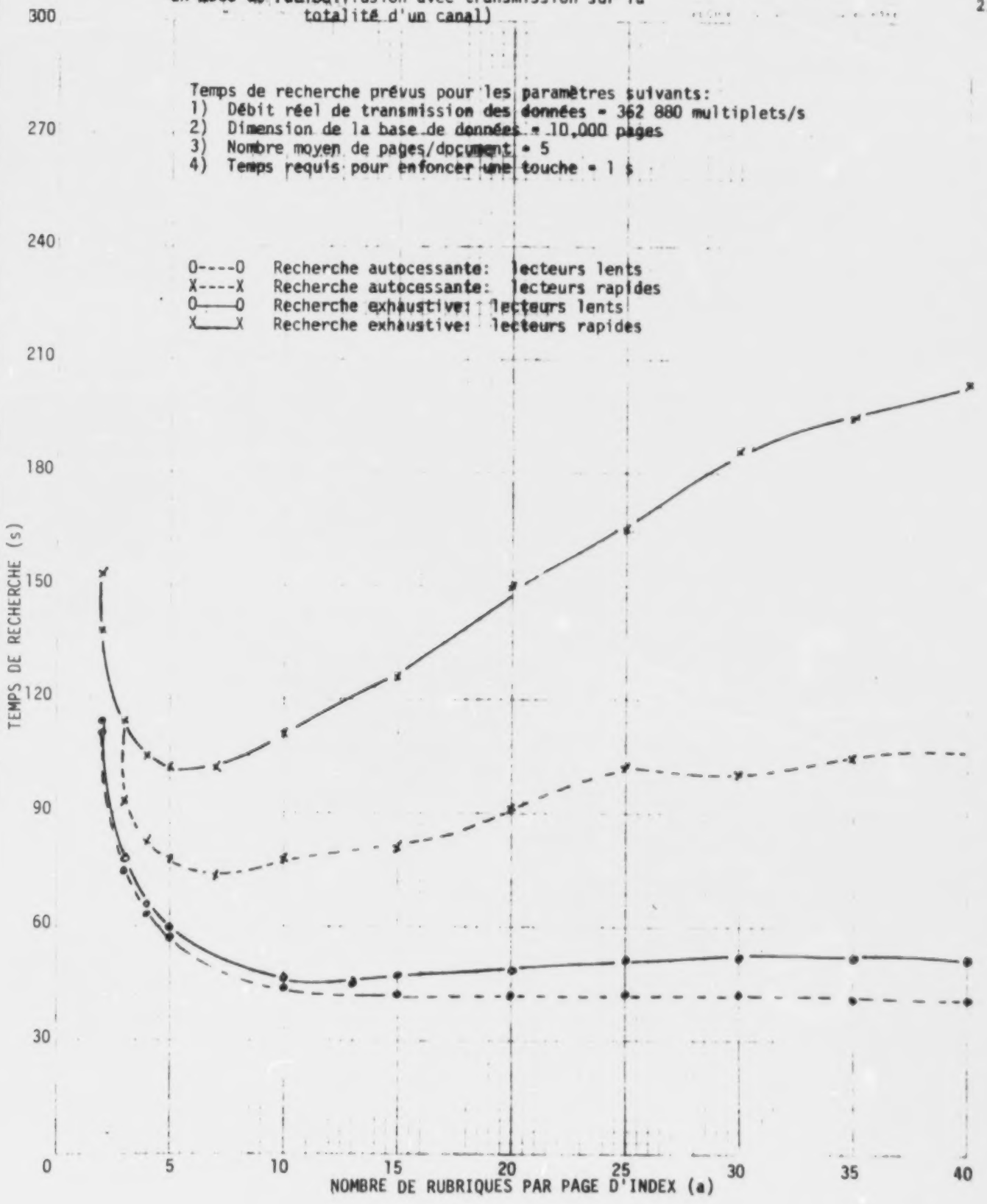


FIGURE 14: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

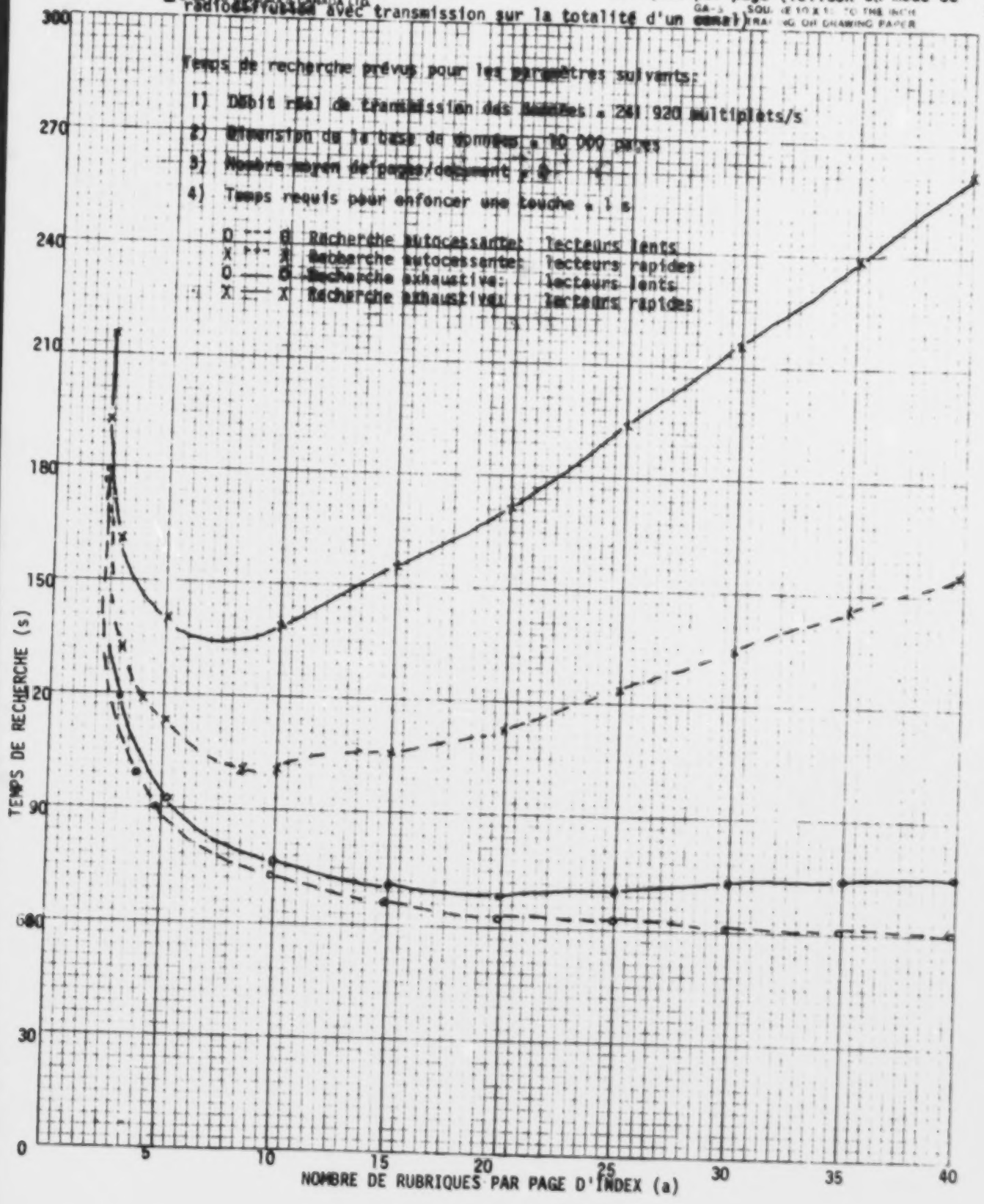


FIGURE 15: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

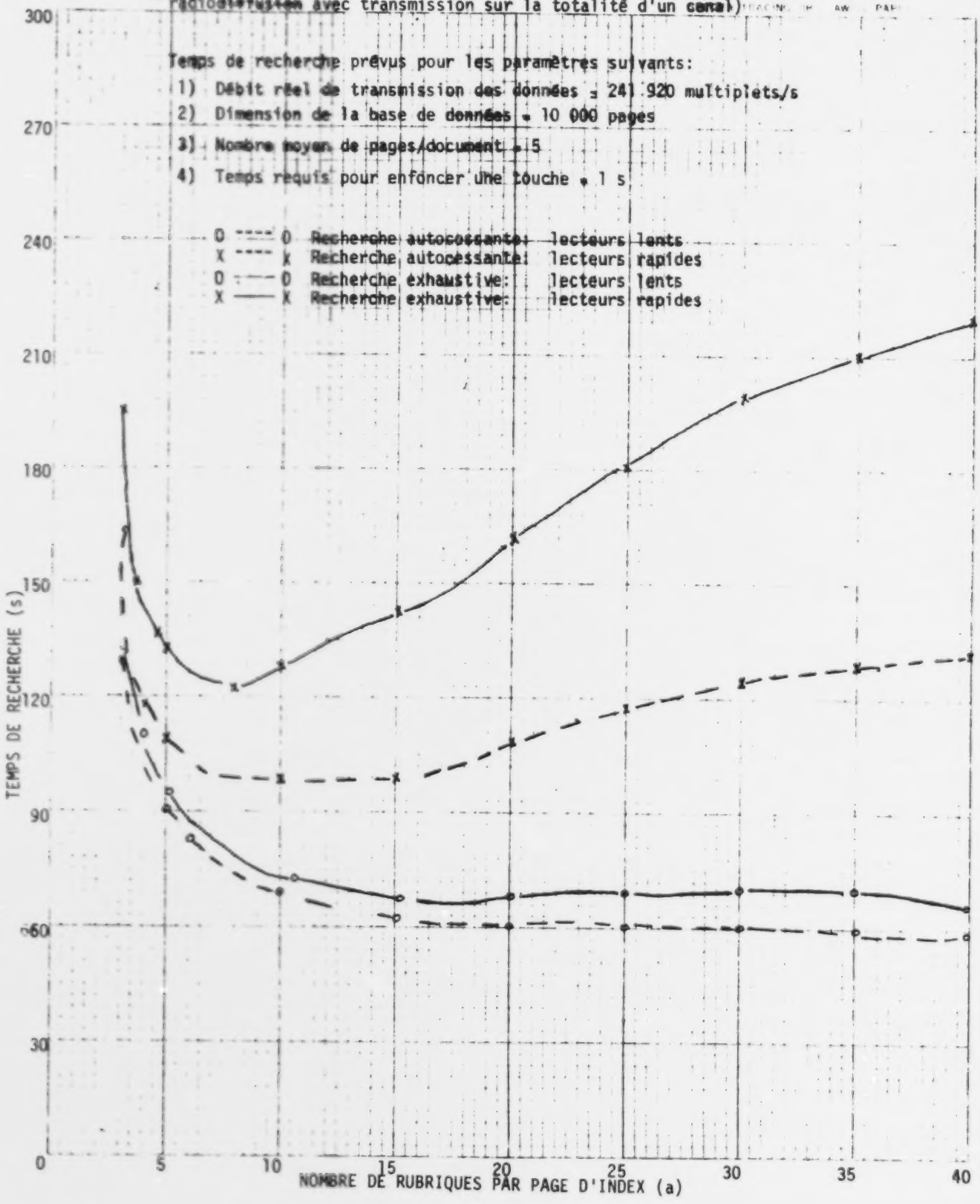


FIGURE 16: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)

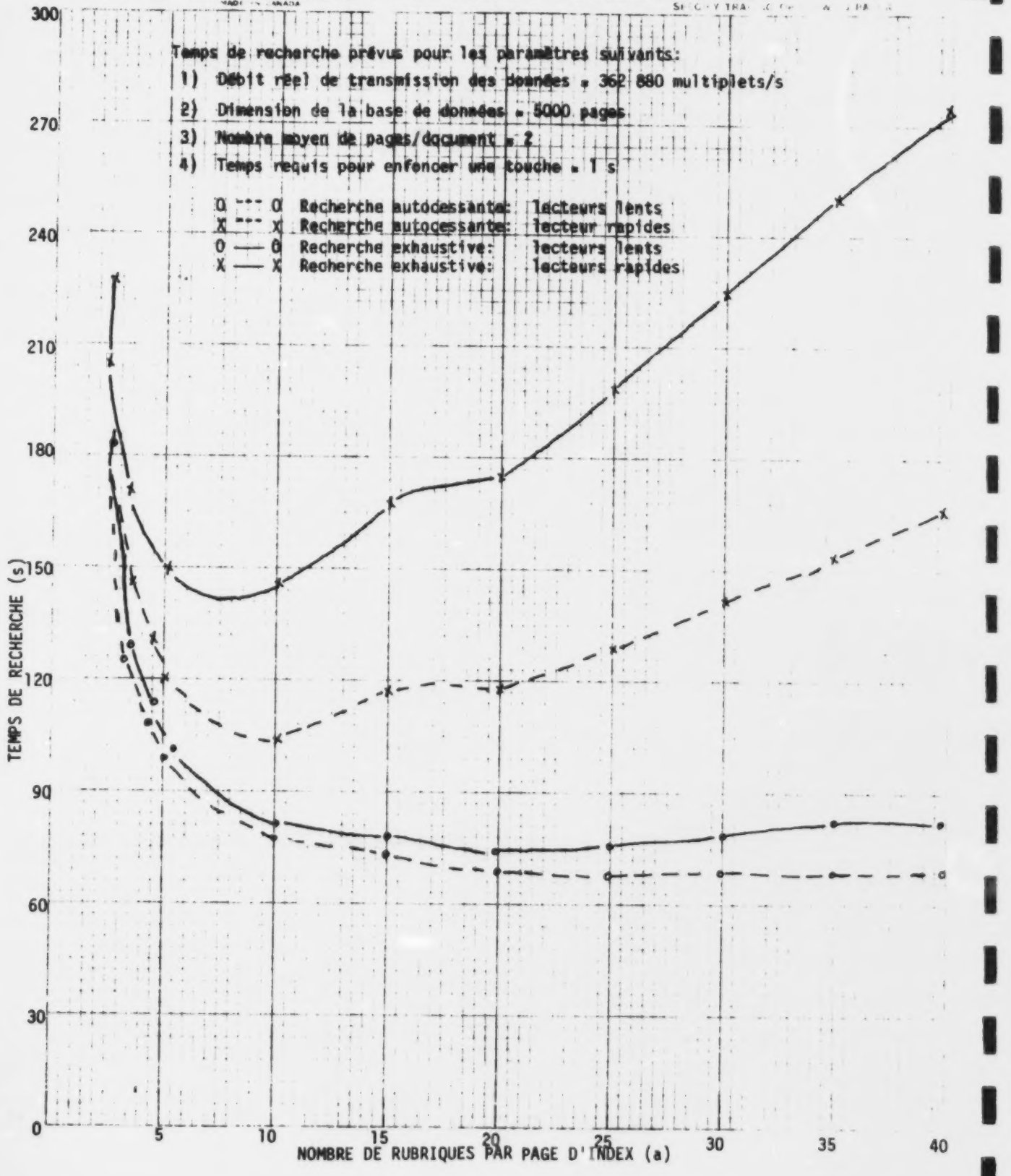
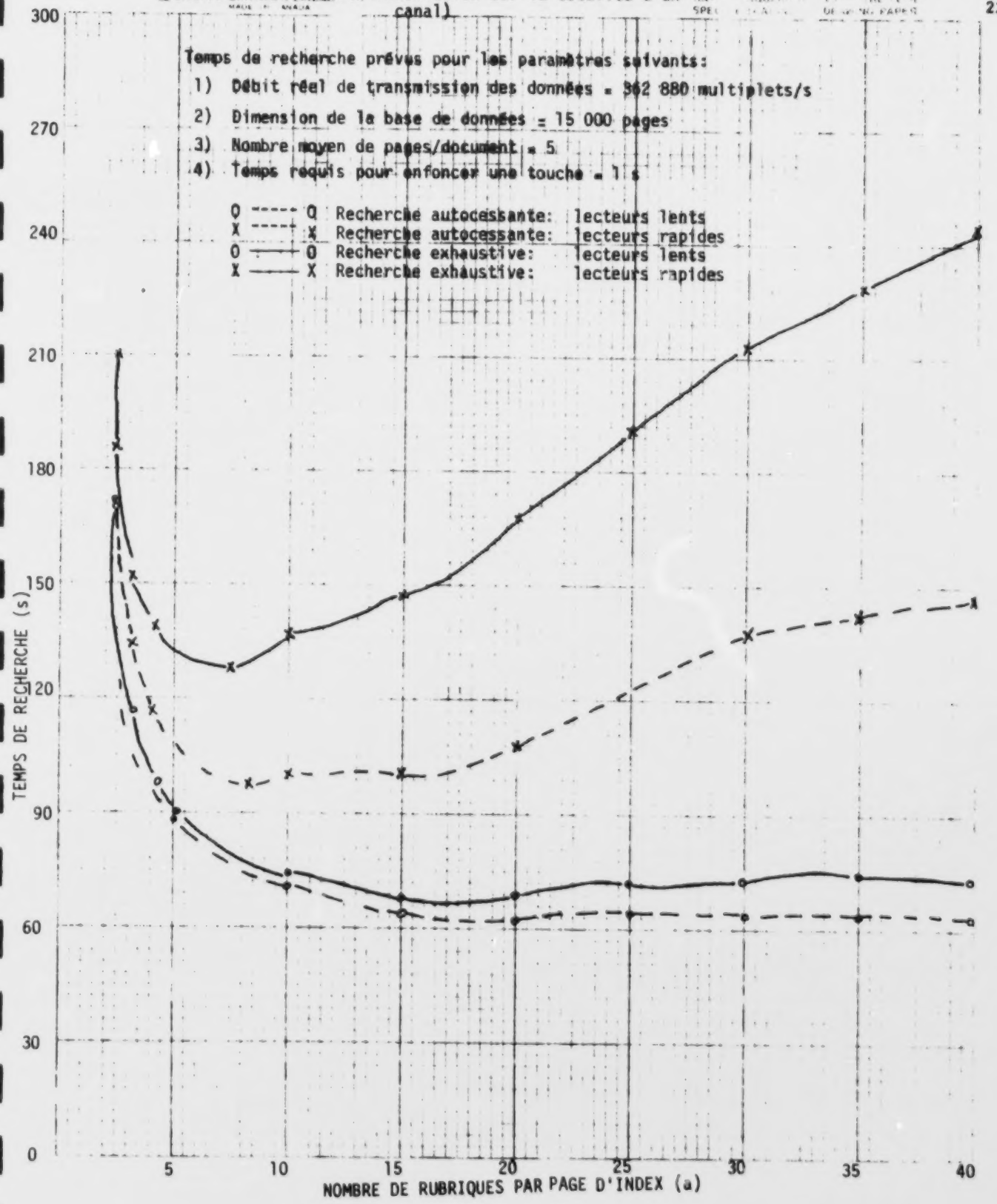


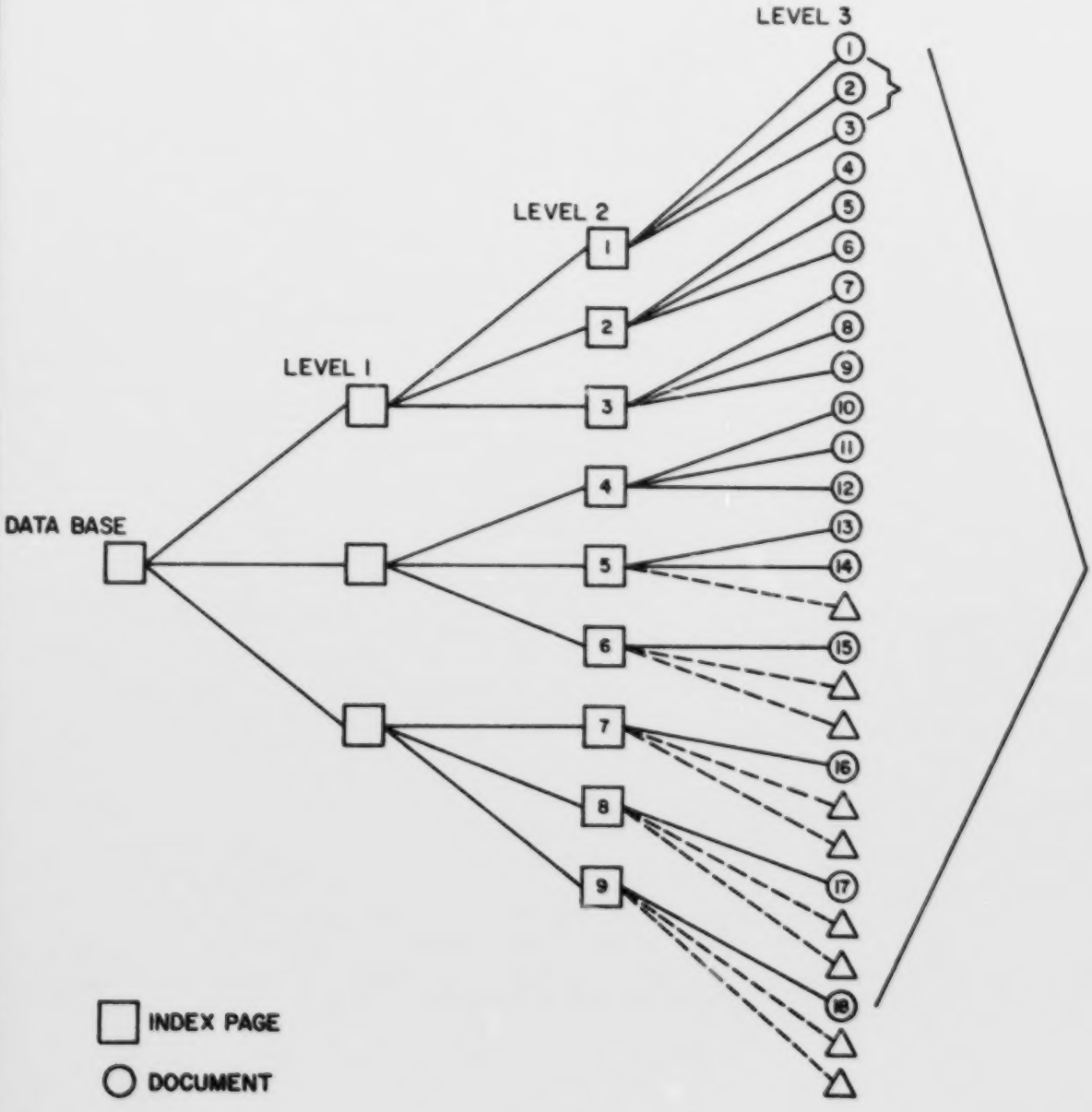
FIGURE 17: Temps de recherche en fonction du nombre de rubriques par page (Télidon en mode de radiodiffusion avec transmission sur la totalité d'un canal)



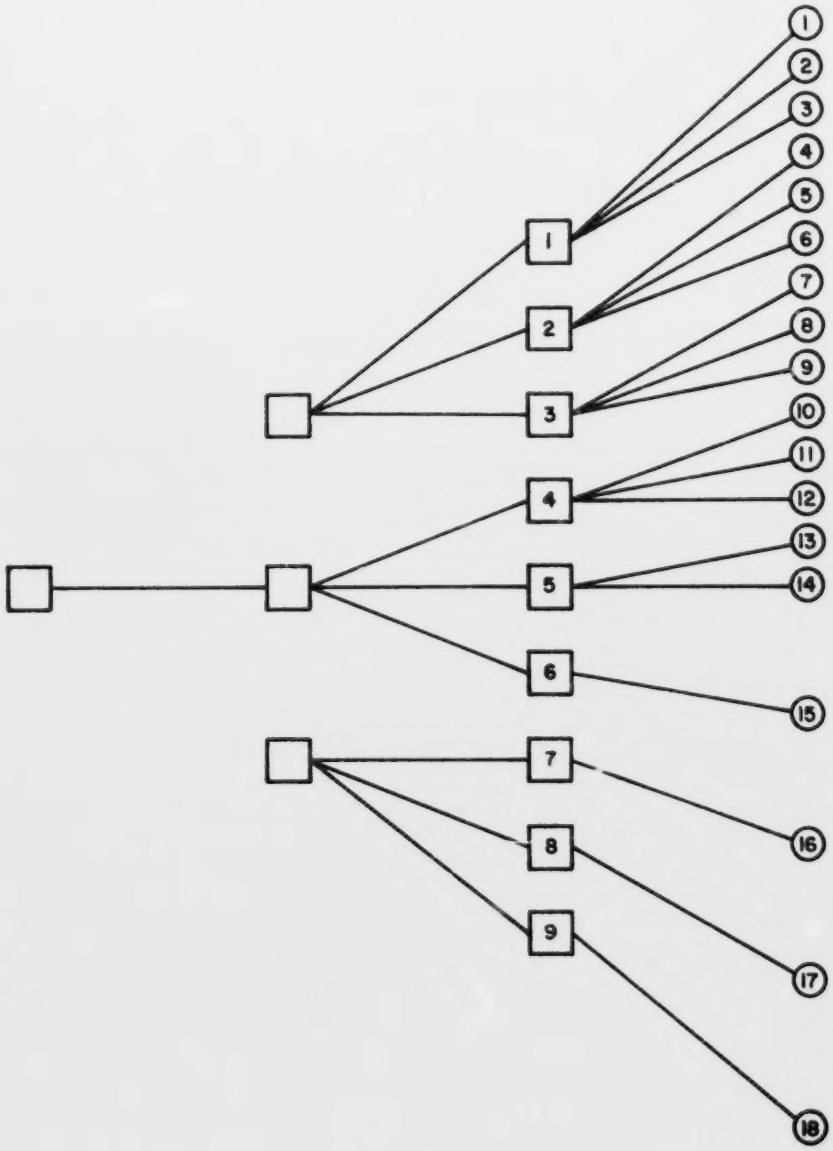
ANNEXE

DISCUSSION SUPPLÉMENTAIRE CONCERNANT L'ORIGINE DE L'ÉQUATION 3

L'exemple ci-dessous clarifie l'origine et l'utilisation de l'équation 3 (voir pages et). Prenons le cas d'une petite base de données contenant 18 documents. L'utilisateur a accès à ces documents au moyen d'un index ramifié comprenant 3 niveaux dont chaque page offre un choix de 3 rubriques. Autrement dit, $d = 18$, $p_m = 3$, $a = 3$. La base de données correspondante est reproduite ci-dessous.



Toutes les pages d'index (ou menus) sont représentées par des carrés, les 18 documents par des cercles et les parcours possibles à l'intérieur de la structure ramifiée, par des lignes. Les triangles représentent les documents qu'on pourrait ajouter à cette base de données sans qu'il soit nécessaire d'augmenter le nombre de rubriques par page d'index ou le nombre de pages d'index. Les pointillés servent à relier ces "documents éventuels" à la structure. On peut constater qu'au moins un des 18 documents (cercles) est relié à chacune des pages d'index au niveau 3 de la structure. En outre, les pages d'index 6 à 9 donnent chacune accès à un des 18 documents (cercles) et à $\underline{a} - 1 = 2$ documents éventuels (triangles). Comme on ne se prévaut jamais de la possibilité d'accès aux documents éventuels, on peutles supprimer de la structure. La figure ci-dessous représente la structure après suppression des triangles.



On peut voir que 4 documents (15 à 18) sont accessibles directement au niveau 3 de la structure. Dans la figure qui précède, la durée moyenne du parcours, $\underline{p}$, est la moins longue. Quand les parcours sont moins longs, les temps de recherche sont plus courts. On peut toujours obtenir cette structure optimale lorsqu'on appaire chaque ensemble de $\underline{a} - 1$ triangles avec un cercle, qu'on supprime les triangles et qu'on élimine les pages d'index qui donnent accès à un seul document (cercle).

L'équation 3 sert à calculer le nombre de documents (cercles) $\underline{n}$ qui sont accessibles au niveau $\underline{p}_m$ (niveau 3 dans notre exemple) de la structure:

$$n = \frac{a^{p_m} - d}{a - 1}$$

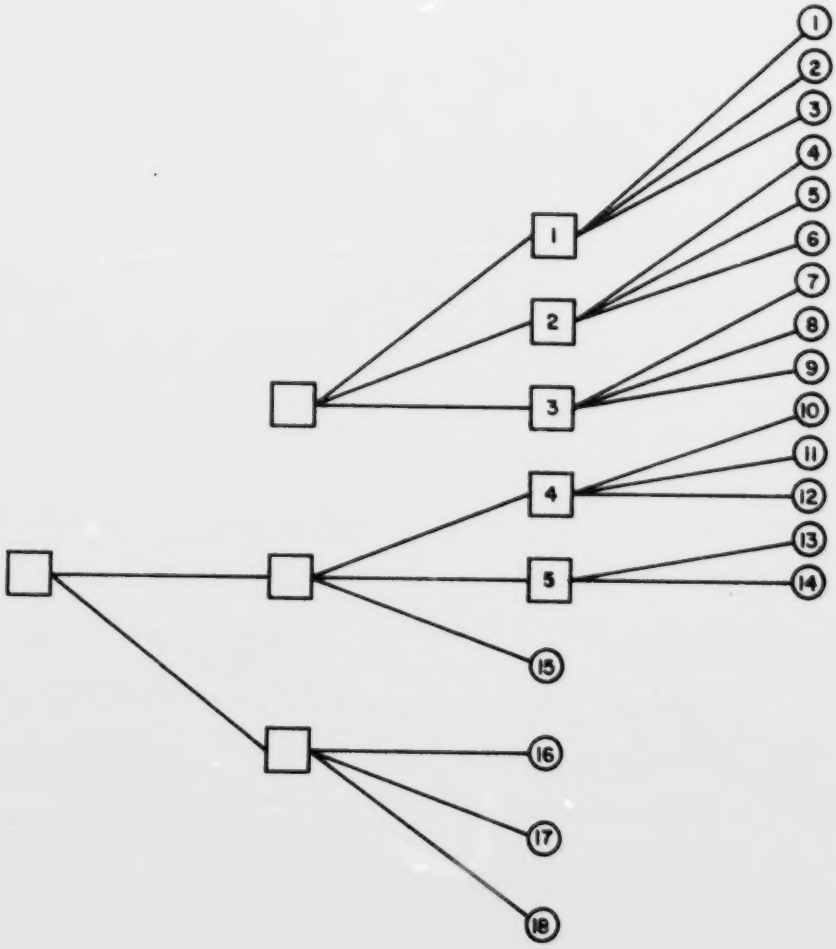
Le numérateur représente ici le nombre de triangles (documents possibles) et le dénominateur, le nombre de triangles qui sont automatiquement supprimés lorsqu'on fait passer le cercle d'un niveau inférieur à un niveau supérieur de la structure. Le terme $\underline{n}$ représente le nombre d'ensembles qu'on peut transformer en appairant un cercle et $\underline{a} - 1$ triangles. Dans notre exemple,

$$n = \frac{2^3 - 18}{3 - 1}$$

$$= 4{,}5 = \underline{\underline{4}}$$

On constatera que le nombre de documents obtenus par l'équation 3 correspond au nombre de documents qui se retrouvent effectivement au niveau 3 de la structure représentée à la figure précédente.

On remarquera aussi que les pages d'index 6 à 9 donnent accès à un seul document (cercle) chacune. Autrement dit, chacune d'elles ne contient qu'une seule rubrique. Elles sont redondantes parce que les documents correspondants (cercles 15 à 18) sont directement accessibles à partir du niveau 2 de la structure comme l'illustre la figure ci-dessous.

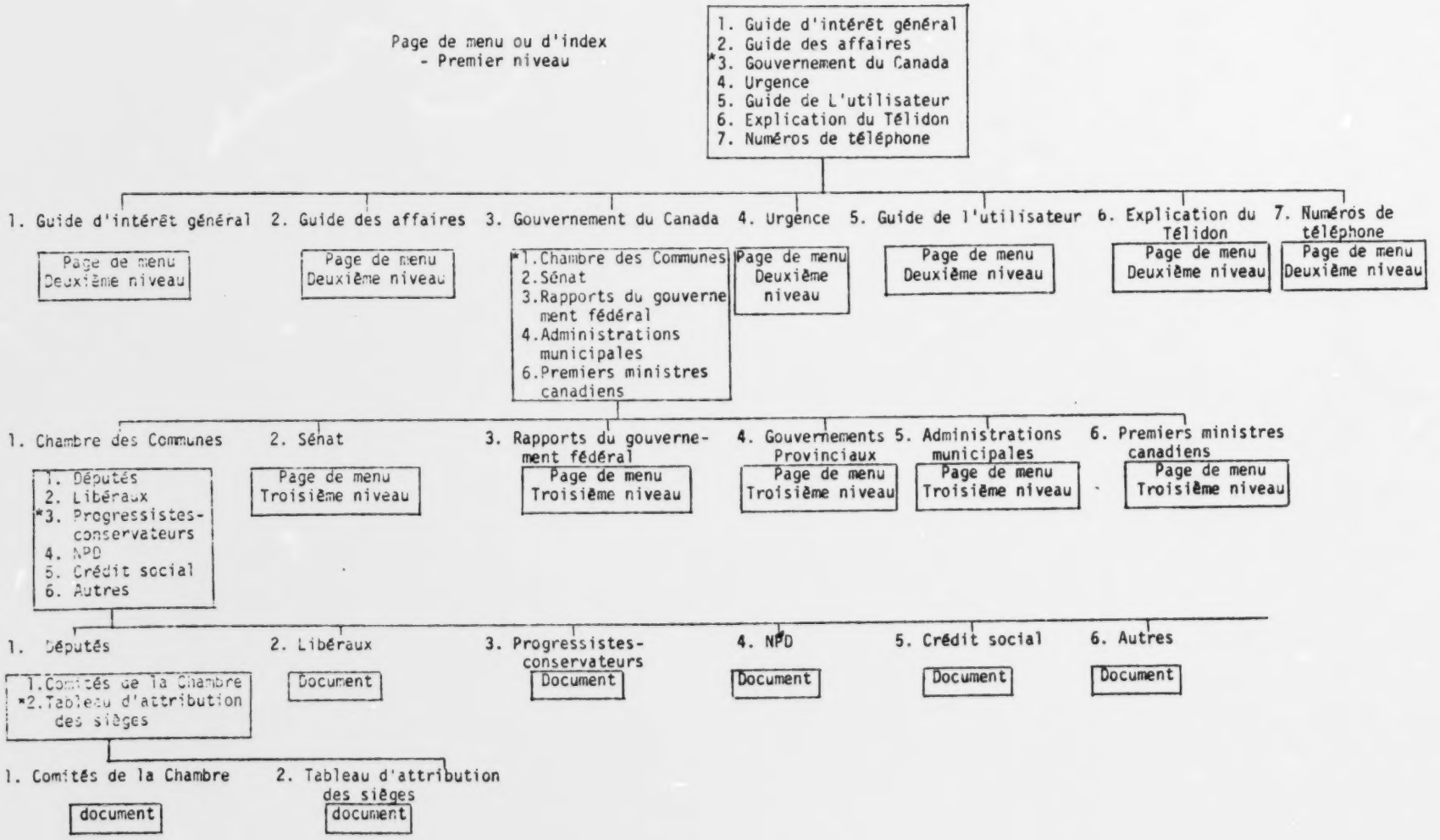


EVALUATION DE LA STRUCTURE ARBORESCENTE DE L'INFORMATION DANS LA BASE DE DONNNES DU TELIDON

Par Eric Lee et Susane Latrémouille
Groupe de recherche en comportement
Ministère des Communications
Ottawa (Ontario)

Le Télidon est une nouvelle application des télécommunications mise au point par une équipe de scientifiques du Centre de recherches du ministère des Communications à Ottawa. Cette innovation technologique permettra de faire chez soi, au moyen du téléviseur, des recherches documentaires dans des bases de données (banques de données), de façon plus souple et plus précise, sur le plan technologique, qu'avec les systèmes publics de recherche documentaire existants. La mise en oeuvre de ce nouveau système amènera probablement une consultation massive de la part du public. C'est pourquoi le mode de communication entre l'ordinateur et l'usager est un facteur essentiel au succès du système. Présentement, les usagers extraient des renseignements du Télidon en faisant des choix dans des menus composés suivant une structure arborescente ou hiérarchique (voir figure 1).

Figure 1: EXEMPLE
D'UNE PARTIE DE LA STRUCTURE
ARBORESCENTE DU TELIDON



La présente étude a pour but d'évaluer une structure arborescente expérimentale en examinant l'aisance avec laquelle les usagers réussissent à trouver des renseignements dans le Télidon. Si l'arbre est bien structuré, ils devraient trouver assez facilement les renseignements recherchés. S'il est mal structuré, la recherche sera laborieuse. Les mots "structure arborescente de l'information" désignent généralement la forme d'organisation des rubriques ou termes d'index servant à obtenir les documents (voir la figure 1). Les pages d'index ou de menu contiennent des listes de termes qui guident le chercheur jusqu'aux documents. La structure comporte un grand nombre de niveaux de rubriques qui se précisent de plus en plus jusqu'au moment où ils permettent d'obtenir le document recherché. Les documents sont les pages qui contiennent le renseignement que cherche l'utilisateur. La structure arborescente sur laquelle porte le présent rapport comprend 900 documents répartis entre 6 niveaux de rubriques ou moins. Chacun des niveaux comporte un menu de 9 rubriques ou moins.

La même information peut être répartie dans un arbre de plusieurs façons différentes. La facilité avec laquelle les usagers trouveront les renseignements dans la base de données pourra varier selon les rubriques que contient cet arbre ou même selon la façon dont elles y sont réparties. La méthode décrite dans notre étude peut servir à l'évaluation de n'importe quelle structure arborescente ou hiérarchique de l'information.

MÉTHODE

Sujets

Dix personnes, dont 6 de l'extérieur du ministère des Communications et 4 attachées au Ministère, ont participé à l'étude. Aucun des participants n'avait utilisé ni même vu le Télidon auparavant.

Épreuve

Lors de l'expérience, la base de données témoin du Télidon contenait 900 documents différents. Chacun des documents contenait une seule catégorie de renseignement. (Par exemple, un document pouvait contenir la liste des films à l'affiche à Ottawa ou les derniers résultats du football ou les inscriptions immobilières, etc.). Pour obtenir un échantillon représentatif des catégories de renseignements susceptibles de figurer dans le Télidon, on a choisi, au hasard, 16 documents différents parmi les 900. Chacune des questions était formulée d'après les renseignements contenus dans le document. Par exemple, la question "Trouvez la carte météorologique de l'endroit où vous êtes présentement" était conçue de façon à susciter la recherche du bulletin météorologique local. Comme on le verra ci-dessous, cela devait amener les participants à faire successivement un certain nombre de choix dans des pages de menu afin de trouver le document. L'article précédé d'un astérisque est celui que devait choisir le chercheur pour trouver le document.

Exemple:

Première page du menu
*1. Intérêt général
2. Guide des affaires
3. Gouvernement du Canada
4. Urgence
5. Guide de l'utilisateur
6. Explication du Télidon
7. Numéros de téléphone

Deuxième page du menu
Guide d'intérêt général
*1. Nouvelles, météo et sports
2. Divertissements

3. Bourse
4. Emploi
5. Voyages
6. Conseils
7. Loisirs
8. Éducation
9. Tableau d'affichage

Troisième page du menu Nouvelles, météo et sports

1. Nouvelles
2. Météo
3. Sports

Quatrième page du menu Météo

*1. Météo locale
2. Météo nationale
3. Bulletins spéciaux

Cinquième page du menu Météo locale

1. Météo locale
2. Prévisions
3. Statistiques
4. Température
*5. Carte

Les questions avaient été formulées avec beaucoup de soin de sorte que les participants n'aient aucun indice quant à l'emplacement des renseignements et chaque demande était suffisamment claire pour que les participants puissent obtenir le renseignement s'il se trouvait dans la banque de données.

Méthode

Chaque participant recevait une série de 16 questions présentées dans un ordre aléatoire différent pour chacun. Pour trouver le document recherché, chaque participant devait choisir des articles dans des menus consécutifs qui lui étaient présentés sur des cartes. Chaque fois qu'un participant faisait un mauvais choix, il devait se reprendre jusqu'à ce qu'il choisisse le bon article. Plusieurs erreurs pouvaient donc porter sur la même page de menu. Chaque fois qu'un participant se trompait, il devait expliquer la raison de son choix et l'expérimentateur enregistrait toutes les erreurs et les raisons. Au total, 79 pages de menu furent présentées à chacun des participants, soit une moyenne d'environ 5 pages de menu par problème. A la fin de la séance, on expliqua aux participants le but de l'étude.

RÉSULTATS ET DISCUSSION

Il n'y a aucun doute que la popularité du Télidon dépendra de la facilité de repérage de l'information dans la base de données. Pour trouver un renseignement, le chercheur devra faire une série de choix. Un mauvais choix le conduira dans la mauvaise section de la banque de données, ce qui augmentera non seulement son temps de recherche mais aussi sa frustration. Plus un chercheur fera d'erreurs, plus il est probable qu'il n'emploiera pas le Télidon.

Les résultats de cette expérience nous fournissent deux types de renseignements: (a) des renseignements sur l'utilisation, c'est-à-dire sur le genre de difficultés auxquelles se heurtent les utilisateurs quand ils cherchent à repérer un élément d'information dans la structure arborescente et (b) des renseignements sur la structure arborescente, c'est-à-dire, sur les endroits de la structure où les utilisateurs éprouvent des difficultés à choisir le bon filon.

Dans l'ensemble, chaque personne s'est trompée en moyenne dans 49,4% (7,9 sur 16) des problèmes essayés. C'est dire que l'usager du Télidon fera probablement une fois sur deux, à peu près, au moins une erreur par consultation. Quatorze des 16 problèmes ont fait l'objet d'au moins un mauvais choix. Le bon choix a été fait à tout coup dans seulement 2 des 16 problèmes.

On peut déduire de la présente expérience que les gens peuvent avoir d'assez grandes difficultés à trouver un élément d'information dans une base de données à structure arborescente ou hiérarchique.

Au total, les 10 utilisateurs ont fait 167 choix incorrects sur un total de 790 présentations de pages de menu. Chacun d'eux a donc fait, en moyenne, 16,7 erreurs dans 16 problèmes. En moyenne, les gens ont fait (16,7/16=1,04) 1,04 erreur chaque fois qu'ils ont consulté la base de données. C'est donc dire que les utilisateurs du Télidon sont susceptibles de faire au moins une erreur chaque fois qu'ils consultent la banque des données à 5 niveaux qui existe présentement.

Dans la présente expérience, 80% de toutes les erreurs ont porté sur seulement 6 des 79 pages de menu et 53% des erreurs ont porté sur les deux premiers niveaux de la structure arborescente. Étant donné que 53% des erreurs ont été faites dans les 2 premiers niveaux et que ce sont les pages du premier niveau qu'on consulte le plus souvent, il serait raisonnable de dire que dans l'ensemble, le public éprouvera de la difficulté à trouver tout renseignement recherché dans cette base de données expérimentale.

Les explications qu'ont fournies les participants sur chacune de leurs erreurs ont beaucoup aidé à comprendre les difficultés survenues au cours de la recherche documentaire. Par exemple, selon certains commentaires, les rubriques de la première page étaient trop générales et l'information devrait être distribuée de façon plus uniforme. Les commentaires relatifs au deuxième

niveau indiquaient un chevauchement entre certaines catégories et quelques termes étaient jugés ambigus.

Souvent, un chercheur faisait plusieurs mauvais choix dans la même page de menu avant de viser juste. Des 790 pages présentées, 679 ont fait l'objet d'un bon choix et 111 présentations ont fait l'objet d'au moins un mauvais choix. Par conséquent, la probabilité d'un bon choix s'établit à 679/790 = 0,86 par page présenté et la probabilité correspondante qu'une personne fasse au moins une erreur chaque fois qu'une page de menu lui est présentée s'établit à 0,14.

Ces résultats ont une incidence sur les structures arborescentes plus vastes. On a éprouvé des difficultés à résoudre la moitié des problèmes soumis alors qu'il suffisait de consulter en moyenne 5 pages de menu. On prévoit que la structure arborescente des bases de données constituées pour le Télidon par les entreprises commerciales auront plus de 15 niveaux (c'est-à-dire que le chercheur devra consulter successivement 15 pages de menu). Avec des bases de données de cette envergure, les gens auront probablement encore de plus grandes difficultés à trouver l'information recherchée. Une base de données à 15 niveaux exigera, par exemple, que le chercheur fasse un bon choix dans 15 pages de menu différentes avant d'obtenir l'information voulue. En comparaison, la base de données actuelle ne demande que 5 bonnes décisions consécutives (c'est-à-dire une bonne décision par page de menu). La probabilité qu'une personne fasse le bon choix dans $\underline{n}$ pages de menu consécutives correspond à la probabilité de faire le bon choix dans une page de menu à la puissance $\underline{n}$. Dans l'hypothèse que la probabilité de faire un bon choix sur chaque page d'une base de données à 15 niveaux corresponde à la probabilité qui s'applique à une base de données à 5 niveaux, (c'est-à-dire que P (probabilité de faire le bon choix dans chaque page de menu)= 0,86), la probabilité de n'éprouver aucune difficulté dans un cas donné s'établit à $(0,86)^{15}$ = 0,10 seulement.

Les gens pourraient donc éprouver des difficultés dans plus de 90% de leurs recherches dans une structure arborescente. Ce taux d'erreur dépasse probablement le seuil de tolérance de la plupart des chercheurs éventuels.

Bien que ces résultats montrent la gravité du problème que pose le repérage de l'information dans une base de données du Télidon, on ne doit pas en conclure que toutes les structures arborescentes (c'est-à-dire les structures hiérarchiques des pages de menu) poseront de graves difficultés à ceux qui les consulteront. En fait, il suffirait de modifier seulement 6 pages de menu des deux premiers niveaux de l'arbre (plus de 80% de toutes les erreurs se sont produites dans 6 pages seulement) pour régler les problèmes relevés dans la banque de données. Les mauvais choix sont dus soit au fait que le renseignement a été classé sous la mauvaise rubrique (terme d'index) dans les pages du menu, soit à l'ambiguïté de la rubrique elle-même. Dans un cas comme dans l'autre, l'arbre peut être modifié de façon à éliminer la plupart de ces mauvais choix. Des recherches plus poussées se poursuivent présentement en vue de montrer comment les fournisseurs d'information peuvent déceler des difficultés dans les pages de menus et comment ces dernières peuvent être modifiées de façon à réduire tant le nombre de mauvais choix que le temps de recherche dans une base de données Télidon. La diminution des erreurs et des temps de recherche rendront sans aucun doute le Télidon plus attrayant.